4 Model Fitting and Uncertainty

David A. Ratkowsky

CONTENTS

4.1 OVERVIEW

This chapter is divided into two main sections, viz. (1) model fitting, featuring the principles of, and examples of, the use of regression models, especially nonlinear regression models, for data sets in food preservation and safety, and comparisons of various modeling approaches and (2) the consequences of uncertainty, i.e., variation in the measurements, and its implications for product shelf life. These sections are designated as 4.2 and 4.3, respectively. The chapter concludes with an epilogue, in which the author raises a few additional issues (Section 4.4).

4.2 MODEL FITTING

This section examines various models used in predictive microbiology, focusing attention upon those criteria and factors that have to be taken into account when the models are being fitted. Also, criteria for assessing goodness-of-fit are presented.

4.2.1 The Models

Three groups of models will be considered in this chapter, which serve to illustrate the various facets of modeling in predictive microbiology. The first group of data (see Table A4.1) and their associated models involve lag time as a function of temperature, and were examined recently by Oscar (2002). The lag time is usually determined experimentally by fitting a primary model, or by noting the time taken before perceptible growth of a bacterial culture is observed. Primary modeling is the subject of Chapter 2 of this book.

Denoting the lag time by λ, the models considered are:
Hyperbola model

$$\lambda = \exp[p/(T - q)] \tag{4.1}$$

Extended hyperbola model

$$\lambda = [p/(T - q)]^m \tag{4.2}$$

Linear Arrhenius model (Davey, 1989)

$$\lambda = \exp[-(A + B/T + C/T^2)] \tag{4.3}$$

Simple square-root model (Ratkowsky et al., 1983)

$$\lambda = 1/\{[b(T - T_{\min})]^2\} \tag{4.4}$$

The second group of data (see Table A4.2) was obtained from experiments conducted by students in the on-going predictive microbiology research program at the University of Tasmania, on the maximum specific growth rate constant μ for three species of microorganism as a function of temperature throughout the entire biokinetic

temperature range. The models for fitting and predicting μ are confined here to just two models, viz. the four-parameter square-root model of Ratkowsky et al. (1983)

$$\sqrt{\mu} = b(T - T_{\min})\{1 - \exp[c(T - T_{\max})]\} \tag{4.5}$$

and the cardinal temperature model of Rosso et al. (1993)

$$\mu = \mu_{opt}(T - T_{\max})(T - T_{\min})^2 / [(T_{opt} - T_{\min})^2(T - T_{opt}) - (T_{opt} - T_{\min})(T_{opt} - T_{\max})(T_{opt} + T_{\min} - 2T)] \tag{4.6}$$

In each of the above models, T represents temperature in degrees absolute, although the only model in which it is essential that degrees absolute be employed is the Linear Arrhenius model (4.3). The other models all involve differences between temperatures, and therefore other temperature scales are acceptable, since they result in equivalent answers. $T_{\min}$ and $T_{\max}$ represent notional minimum and maximum temperatures, respectively, the term "notional" meaning that these temperatures are not to be interpreted as "true" minimum and maximum temperatures, although various authors have mistakenly or misguidedly given them this interpretation (e.g., Dantigny and Molin, 2000). In 4.5 and 4.6, they are nothing more than intercepts on the rate (μ) axis, i.e., the temperatures at which the rate equals zero when a graph of μ vs. T is extrapolated outside the range of the observed data. In 4.6, the additional cardinal temperature T_{opt} represents the temperature at which growth is optimal (i.e., μ is greatest), and the fourth parameter μ_{opt} is the maximum specific growth rate corresponding to T_{opt}. Thus, the cardinal temperature model (4.6) is the only one in which all its parameters can be considered to be biologically interpretable, although not necessarily achievable (i.e., $T_{\min}$ and $T_{\max}$). All other models contain arbitrary constants, viz. *p, q, m, A, B, C, b,* and *c*, which are devoid of biological meaning. They are simply parameters included in the model to enhance the curve-fitting prospects of the model.

It should also be noted that 4.5 and 4.6 apply only in the range $T_{\min} < T < T_{\max}$, and that outside these ranges, i.e., for $T < T_{\min}$ and $T > T_{\max}$, the rate μ is zero. To be mathematically correct, these bounds should be stated along with the equation definitions, but they are omitted here for simplicity, and it should be understood that nonzero rates can only apply at temperatures between $T_{\min}$ and $T_{\max}$. It should also be self-evident that when modeling data, using either of the above models, only nonzero rates should be employed. Data corresponding to temperatures at which the observed rates are zero need to be discarded when curve fitting.

4.2.2 Stochastic Assumptions

Equation 4.1 to Equation 4.6 are nonlinear regression models with two to four parameters, which may be estimated using nonlinear regression modeling. Some of the equations can be transformed by rearranging terms, thereby linearizing them.

For example, the reason why 4.3 is called a “Linear Arrhenius” model can be seen by rewriting it as:

$$\ln \text{rate} = \ln(1/\lambda) = A + B/T + C/T^2$$

The result follows from the fact that a time, whether it is a lag time, a generation time, or some other time-based variable, may be expressed as a rate by taking its reciprocal. The right-hand side is a quadratic polynomial in $1/T$, the reciprocal of absolute temperature, a term that is often seen in Arrhenius-type models. Similarly, Equation 4.4 may be rearranged as:

$$\sqrt{\text{rate}} = \sqrt{(1/\lambda)} = b(T - T_{\text{min}})$$

which shows that the model is in reality the simple square-root model of Ratkowsky et al. (1982). Whether one should or should not transform response variables in this manner is decided by the so-called *stochastic assumption*, i.e., the assumption that one makes about how the response variable, the lag time λ or the specific growth rate constant μ, varies with change in the explanatory variable, the temperature T. The lag time λ will almost never have a homogeneous variance, as the lag time in the suboptimal range tends to be much more variable at low temperatures where growth rates are slow, than near the optimal temperature T_{opt}. Therefore, for modeling 4.1 to 4.6, careful consideration has to be given to the form in which these models are fitted, to reflect the stochastic assumption made.

For the specific growth rate constant μ, Ratkowsky et al. (1983), in the paper in which the four-parameter square-root model (4.5) first appeared, assumed that the variance was homogeneous in $\sqrt{\mu}$*;* that is, the transformed response $\sqrt{\mu}$ was assumed to have the same variance at each temperature T. This implies that the variance of the untransformed μ is a function of T, the variance increasing as μ increases. The near constancy of the variance of $\sqrt{\mu}$ has previously been demonstrated by R.K. Lowry (unpublished data) on a variety of data sets when the square-root model was first developed for suboptimal data sets (Ratkowsky et al., 1982).

On the other hand, Rosso et al. (1993) implicitly assumed that the variance of μ was homogeneous (i.e., unchanging with T). This assumption results in a different set of parameter estimates from what is obtained by assuming that $\sqrt{\mu}$ is homogeneous in T. An alternative assumption, also frequently used in the predictive microbiology literature (e.g., see Schaffner, 1998), is that $\ln \mu$ is homogeneous in μ, where $\ln \mu$ is the natural logarithm of the rate constant μ. (One may use “base 10” logarithms but mathematicians prefer the “base *e*” natural logarithms.) Incorporation of the stochastic assumption is most easily done by applying the transformation to both the left-hand side and the right-hand side of the expression. The result is a proliferation of forms in which the same basic equation may appear, each of which depends upon the stochastic assumption. The equations that follow express the other alternative forms in which the models used in this chapter may appear.

Hyperbola model

Log assumption

$$\ln(1/\lambda) = -p/(T - q) \tag{4.1a}$$

Square-root assumption

$$\sqrt{(1/\lambda)} = \sqrt{\exp[-p/(T-q)]} \tag{4.1b}$$

Extended hyperbola model

Log assumption

$$\ln(1/\lambda) = -m \ln p + m \ln(T - q) \tag{4.2a}$$

Square-root assumption

$$\sqrt{(1/\lambda)} = [p/(T-q)]^{-m/2} \tag{4.2b}$$

Linear Arrhenius model

Log assumption

$$\ln(1/\lambda) = A + B/T + C/T^2 \tag{4.3a}$$

Square-root assumption

$$\sqrt{(1/\lambda)} = \sqrt{\exp(A + B/T + C/T^2)} \tag{4.3b}$$

Simple square-root model

Square-root assumption

$$\sqrt{(1/\lambda)} = b(T - T_{\text{min}}) \tag{4.4a}$$

Log assumption

$$\ln(1/\lambda) = 2 \ln b + 2 \ln(T - T_{\text{min}}) \tag{4.4b}$$

Four-parameter square-root model

Rate assumption

$$\mu = b^2(T - T_{\text{min}})^2\{1 - \exp[c(T - T_{\text{max}})]\}^2 \tag{4.5a}$$

Log assumption

$$\ln \mu = 2\log b + 2\log(T - T_{\min}) + 2\log\{1 - \exp[c(T - T_{\max})]\} \tag{4.5b}$$

Cardinal temperature model

Square-root assumption

$$\sqrt{\mu} = \sqrt{\{\mu_{opt}(T - T_{\max})(T - T_{\min})^2 / [(T_{opt} - T_{\min})^2(T - T_{opt}) - (T_{opt} - T_{\min})(T_{opt} - T_{\max})(T_{opt} + T_{\min} - 2T)]\}} \tag{4.6a}$$

Log assumption

$$\ln \mu = \ln \mu_{opt} + \ln(T - T_{\max}) + 2\ln(T - T_{\min}) - \ln[(T_{opt} - T_{\min})^2(T - T_{opt}) - (T_{opt} - T_{\min})(T_{opt} - T_{\max})(T_{opt} + T_{\min} - 2T)] \tag{4.6b}$$

4.2.3 Data Sets and Software

To illustrate the methodology, and show the effects of the various stochastic assumptions, several groups of data are employed. The first group is given in Table A4.1, and involves the lag time for the growth of *Salmonella typhimurium* on autoclaved, and therefore sterile, ground chicken breast and thigh burgers at 2°C intervals from 8 to 48°C. Colonies were counted on inverted spiral plates after incubation at 30°C for 24 h. The two sets of data (breast vs. thigh) enable a comparison of regressions to be made to test whether the same model can successfully fit both data sets.

The second group of data is for the specific growth rate constant μ vs. temperature, these being the same three data sets as used by Lowry and Ratkowsky (1983). They involve an *Alteromonas* sp. (CLD38), the temperatures ranging between 1.3 and 29.9°C, a *Pseudomonas* Group I species (16L16) in the range of 0 to 31.6°C, and a mesophilic species *Morganella morganii* (M68) (formerly *Proteus morganii)*, with data in the range of 19 to 41.5°C. Unlike the data from Oscar (2002), turbidimetric measurements, rather than plate counts, were used. Table A4.2 lists these data sets, which are used to compare the four-parameter square-root and cardinal temperature models (4.5 and 4.6, respectively).

The third group of data is for the growth of *Listeria monocytogenes* and involves five complete replicates of growth data throughout the entire biokinetic range for temperature. Four of these replicates were used in a recently published study on variation of branched-chain fatty acids (Nichols et al., 2002), with a fifth set of data being added, which was not used in that study because it lacked fatty acid compositions. Different temperatures were obtained using a temperature gradient incubator and growth was monitored by measuring the percentage of transmittance of light at a wavelength of 540 nm. The data are given in Table A4.3, and are expressed as the square root of rate vs. temperature in degree Celsius. Note that there are a few zero rates at some low and some high temperatures. As indicated in Section 4.2.1, these

data points have to be discarded before modeling can begin. These data sets will be fitted using the square-root model (4.5) and are used here to illustrate methodology for the examination of residuals and the effect of replication.

Nonlinear regression modeling was carried out using the SAS© statistical software, Version 8.2, PROC NLIN. The Gauss–Newton method was chosen as the fitting option. No derivatives need to be supplied, as the procedure computes them automatically. A measure of nonlinear behavior of the parameter estimators, the Hougaard measure of skewness (see Ratkowsky, 1990, pp. 27–28), was calculated using the option Hougaard. Even when the regression model was linear, e.g., 4.3a, PROC NLIN was still employed, as the Gauss–Newton method converges to the correct least-squares estimates in a single iteration, irrespective of the initial parameter values.

4.2.4 Results of Model Fitting

4.2.4.1 Lag Time Modeling

Oscar (2002) concluded that lag times were similar for breast and thigh meat for all temperatures in the data set of Table A4.1, probably as a consequence of the autoclaving process, and combined the individual data sets (n = 21 data points for each) into a single data set (n = 42). The graphs shown in Figure 4.1a (time scale) and Figure 4.1b (ln rate scale) visually indicate that the data sets are similar, and this will be confirmed by formal testing later in this chapter (see Section 4.2.5.2). We will use the combined data set to test the efficacy of the models 4.1 to 4.4 in the paragraphs that follow.

Table 4.1 presents parameter estimates and their asymptotic standard errors obtained by fitting models 4.1 to 4.4 to the data in Table A4.1 using the lag time λ as the response variable, and also from models 4.1a to 4.4b, which incorporate the logarithmic and the square-root transformations, respectively, applied after converting lag time into rate by taking the reciprocal of λ. Table 4.2 presents the residual mean squares corresponding to these models.

Superficially, from Table 4.2 it appears that the extended hyperbola model (4.2) is best, having a smaller residual mean square than the three alternative models when lag time λ is used as the response variable and also using the "log rate" stochastic assumption, but the Linear Arrhenius model (4.3) has a slightly smaller error mean square using the "square root of rate" assumption. Irrespective of assumption, the hyperbola model (4.1) performs badly and the simple square-root model (4.4) is intermediate. The parameter estimates given in Table 4.1 show a strong dependence upon the stochastic assumption for all models. The reasons for this will be explored in subsequent sections of this chapter.

4.2.4.2 Modeling μ

Table 4.3 lists the parameter estimates obtained from models 4.5 and 4.6 for the data in Table A4.2 using the rate assumption i.e., with μ as the response variable, and also with the square-root assumption (i.e., with $\sqrt{\mu}$ as the response) and the log assumption (i.e., with ln μ as the response). Table 4.4 presents the residual mean squares corresponding to these models.

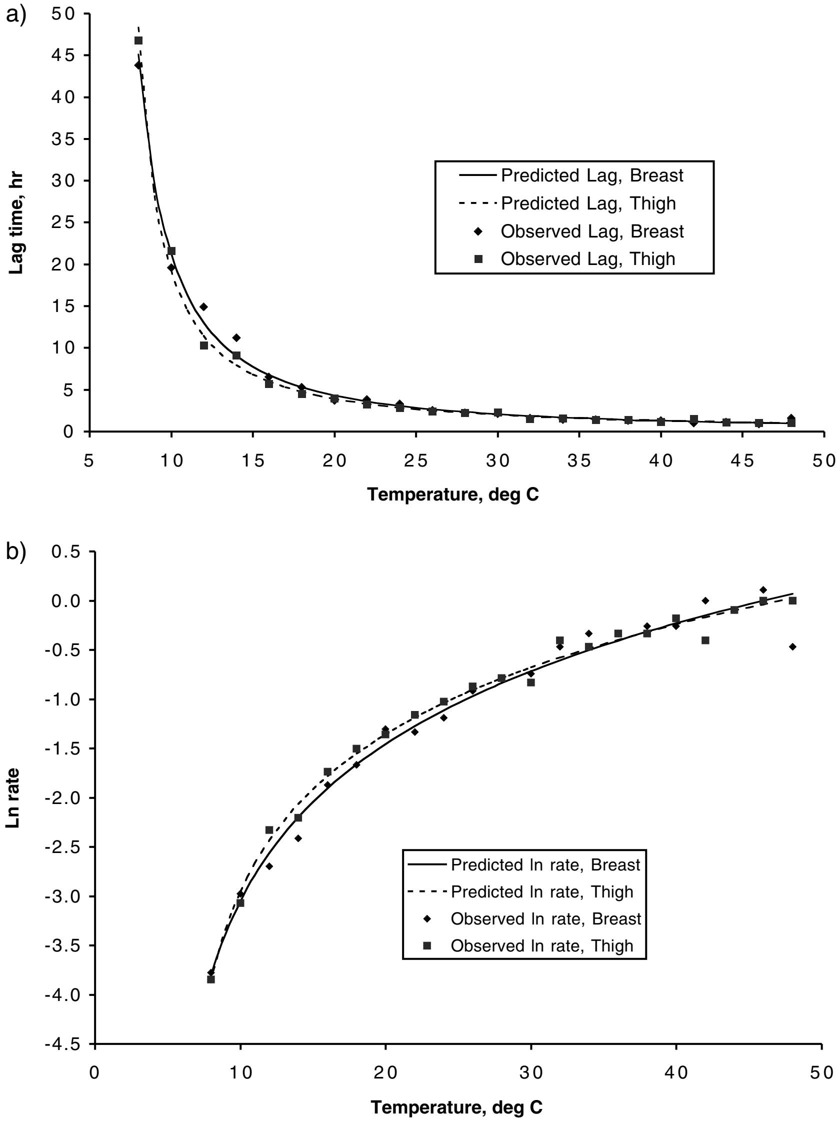


FIGURE 4.1 (a) Observed and predicted lag times vs. temperature. (b) Observed and predicted ln rates vs. temperature. Predicted values were obtained using the extended hyperbola model (4.2) and the log rate assumption on the breast and thigh data separately.

The parameter estimates in Table 4.3 indicate that they are not strongly dependent upon the stochastic assumption. For example, the maximum range of the estimates of T_{min} or T_{max} from either model is 1.2 degree and is less than half a degree for T_{opt} in the cardinal temperature model. We have seen from the results for lag time modeling in Table 4.1 that the stochastic assumption in regression modeling can have a big impact on the magnitude and the precision of the estimates. That it

TABLE 4.1
Parameter Estimates and Their Asymptotic Standard Errors for Models 4.1 to 4.4 in Their Original and Transformed Forms, Data of Table A4.1 for the Different Stochastic Assumptions

		Assumption		
Model	**Parameter**	**Untransformed**	**Log Rate**	**Square-Root Rate**
(4.1) Hyperbola	p	28.9 ± 1.02	20.9 ± 1.51	15.5 ± 1.51
	q	0.434 ± 0.278	2.846 ± 0.516	5.69 ± 0.89
(4.2) Extended hyperbola	p	40.6 ± 2.81	41.0 ± 0.96	41.1 ± 1.11
	q	5.23 ± 0.32	5.66 ± 0.41	6.48 ± 0.87
	m	1.42 ± 0.09	1.34 ± 0.07	1.23 ± 0.09
(4.3) Linear Arrhenius	A	–540.5 ± 37.1	–274.5 ± 19.8	–214.2 ± 20.7
	B	3.52 ± 0.25	1.74 ± 0.13	1.34 ± 0.14
	C	–0.0057 ± 0.0004	–0.0028 ± 0.0002	–0.0021 ± 0.0002
(4.4) Simple square-root	B	0.0312 ± 0.0009	0.0236 ± 0.0007	0.0207 ± 0.0008
	T_{min}	3.22 ± 0.16	0.60 ± 0.51	–2.93 ± 1.27

TABLE 4.2
Residual Mean Squares for the Four Models 4.1 to 4.4 in Their Original and Transformed Forms, Data of Table A4.1 for the Different Stochastic Assumptions

	Assumption		
Model	**Untransformed**	**Log Rate**	**Square-Root Rate**
(4.1) Hyperbola	1.0896	0.0855	0.00910
(4.2) Extended hyperbola	0.6787	0.0185	0.00251
(4.3) Linear Arrhenius (T in Kelvin)	2.4674	0.0343	0.00245
(4.4) Simple square-root	1.0090	0.0394	0.00388

has not for the rate data is probably a reflection of the fact that the data fit each of the models well, as evidenced by the low residual mean squares in Table 4.4. The better the fit of the data to the model, the lesser the importance of the stochastic assumption. In the limit, a perfect fit would result in a fitted model that is independent of the error assumption.

Other differences may be observed from the examination of the estimates. Estimates of T_{min} from the square-root model are consistently lower than those from the cardinal temperature model, whereas estimates of T_{max} from the square-root model are consistently higher than those from the cardinal temperature model. This means that the predicted "biokinetic range," the temperature range at which nonzero growth is predicted, is always higher when estimated from the square-root model than when estimated from the cardinal temperature model. Despite the differences

TABLE 4.3
Parameter Estimates from Models 4.5 and 4.6 Fitted to the Data of Table A4.2 for the Different Stochastic Assumptions

Model	Parameter	Assumption: Square-Root Rate	Rate	Log Rate
CLD38 (Temperature Estimates in Kelvin)				
(4.5) 4-Parameter square root	T_{min}	266.9	267.2	266.7
	T_{max}	309.3	309.5	309.0
	b	0.0100	0.0102	0.0099
	c	0.1817	0.1732	0.1929
(4.6) Cardinal temperature	T_{min}	267.6	268.1	267.2
	T_{max}	306.2	306.5	305.8
	T_{opt}	299.1	299.0	299.3
	μ_{opt}	0.0742	0.0738	0.0747
16L16 (Temperature Estimates in Kelvin)				
(4.5) 4-Parameter square root	T_{min}	266.2	266.6	265.9
	T_{max}	310.4	310.9	309.7
	b	0.0107	0.0110	0.0105
	c	0.3096	0.2773	0.3572
(4.6) Cardinal temperature	T_{min}	266.6	267.3	266.1
	T_{max}	306.9	307.5	306.3
	T_{opt}	302.7	302.5	302.8
	μ_{opt}	0.1274	0.1263	0.1293
M68 (Temperature Estimates in Kelvin)				
(4.5) 4-Parameter square root	T_{min}	272.1	272.0	272.1
	T_{max}	317.5	317.6	317.5
	b	0.00227	0.00227	0.00227
	c	0.3397	0.3390	0.3420
(4.6) Cardinal temperature	T_{min}	274.8	275.1	274.4
	T_{max}	315.9	316.0	315.9
	T_{opt}	310.0	310.0	310.1
	μ_{opt}	0.00623	0.00624	0.00622

when these models are extrapolated, both models closely fit the data within the observed range of the data (see Figure 4.2a to Figure 4.2c). There does not appear to be any systematic departure of either model from the data. From the residual mean squares from each model in Table 4.4, it can be seen that the square-root model fits better, regardless of the stochastic assumption, for two of the three data sets, but the cardinal temperature model fits slightly better, depending upon the stochastic assumption, for the third data set. More data sets are needed to see if there is a consistent pattern. These limited results suggest that there is little to choose between the two models in terms of their ability to fit data over the whole of the temperature range. Further examination of goodness-of-fit will be made in the following sections.

TABLE 4.4
Residual Mean Squares for Models 4.5 and 4.6 Fitted to the Data of Table A.2 for the Different Stochastic Assumptions

Model	Assumption		
	Square-Root Rate	Rate	Log Rate
	CLD38		
(4.5) 4-Parameter square root	5.924×10^{-6}	0.735×10^{-6}	9.82×10^{-4}
(4.6) Cardinal temperature	8.536×10^{-6}	1.091×10^{-6}	13.0×10^{-4}
	16L16		
(4.5) 4-Parameter square root	8.75×10^{-6}	1.73×10^{-6}	10.6×10^{-4}
(4.6) Cardinal temperature	13.5×10^{-6}	2.69×10^{-6}	14.1×10^{-4}
	M68		
(4.5) 4-Parameter square root	1.17×10^{-6}	2.35×10^{-8}	10.4×10^{-4}
(4.6) Cardinal temperature	1.16×10^{-6}	2.14×10^{-8}	11.2×10^{-4}

4.2.5 Measures of Goodness-of-Fit

4.2.5.1 R^2 and "Adjusted R^2" Are Not Appropriate in Nonlinear Regression

The residual mean squares presented in Table 4.2 to Table 4.4 are part of the process of assessing the goodness-of-fit of a regression model. An oft-used criterion appearing in the scientific literature for judging whether or not a regression model fits well is R^2, the proportion of "explained variation" based upon the sum of squares; that is, it is nothing more than the ratio of the sum of squares due to regression to that of the total sum of squares of the response variable around its mean. As such, it purports to indicate how much of the total variation in the response variable, ignoring the regression, is explained by the regression model, i.e., by the inclusion of terms which are introduced to help explain the variation in the response variable. Similarly, another goodness-of-fit measure, the so-called "adjusted R^2" or "percent variance accounted for," is based upon the variances (i.e., the mean squares) rather than upon the sum of squares. The use of either of these measures for nonlinear regression is inappropriate, usually leading to a rather overoptimistic view of the success of the modeling process. We now look into these measures, and some alternatives to them, in some detail.

In linear regression models, where the model contains an intercept, as in the simple straight-line model

$$Y = \alpha + \beta X \tag{4.7}$$

where X is an explanatory variable and Y a response variable, or in the multiple regression model

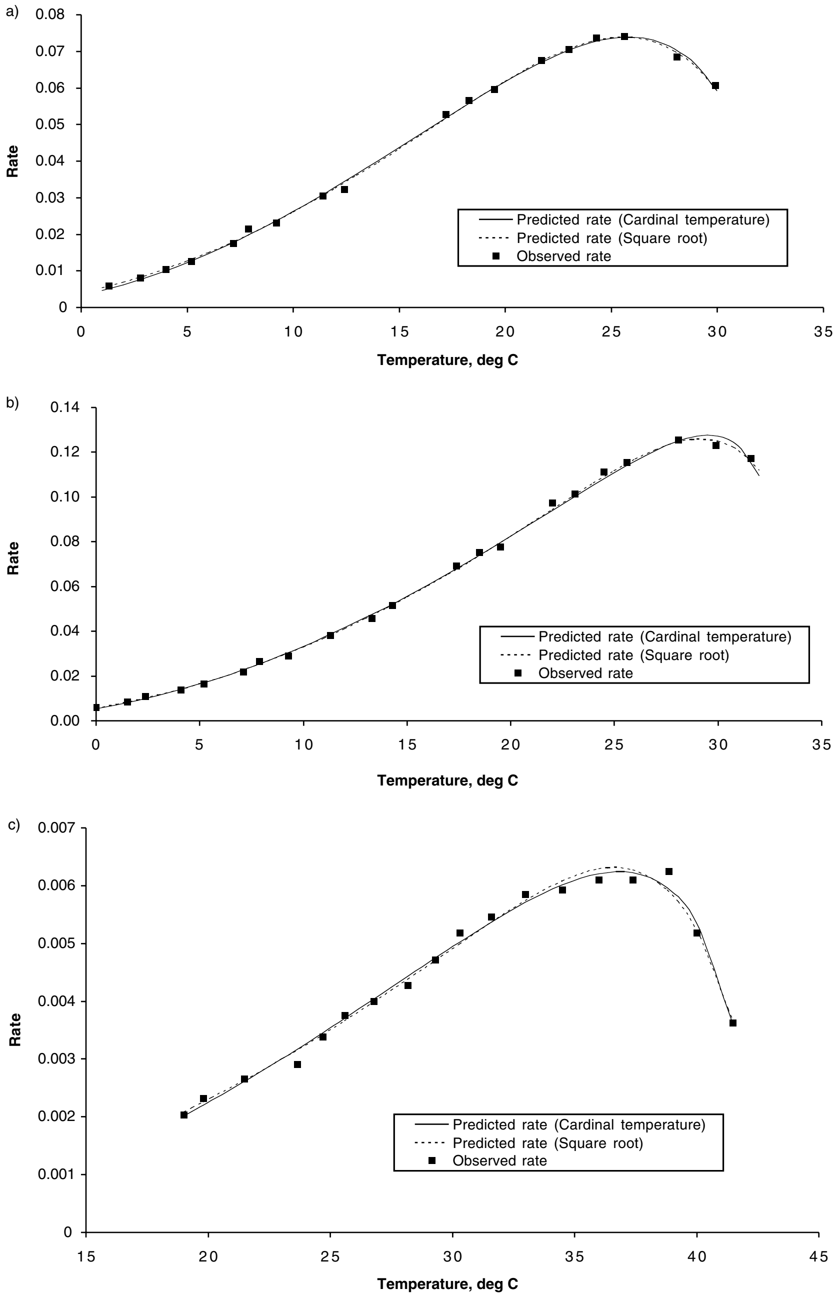


FIGURE 4.2 Observed and predicted rates vs. temperature for (a) CLD38 data, (b) 16L16 data, and (c) M68 data. Predicted rates were obtained using the cardinal temperature and square-root models.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_p X_p \tag{4.8}$$

where $X_1, X_2, \ldots, X_p$ are explanatory variables and Y a response variable, the parameters $\alpha, \beta, \beta_1, \beta_2, \ldots, \beta_p$ are estimated using the criterion of least squares. This widely used criterion finds a set of parameter estimates that minimizes the sum of squares of the differences between the observed and the fitted points, these differences being referred to as the residuals. The ratio of the sum of squares of the residuals to the corrected sum of squares of the response variable Y (the denominator being the sum of squares of the observed Ys around its mean) is the complement of R^2, also known as the "coefficient of determination," being the proportion of the total variation of Y (about its mean) that is explained by the regression.

For a linear regression model with an intercept (e.g., A in Equation 4.3a), the use of R^2 as a measure of goodness-of-fit seems sensible, but even there it may be misleading. As pointed out by Sen and Srivastava (1990, p.14), R^2 depends not only on the sum of squares of the residuals, as one would wish, but also on the corrected sum of squares of the response variable about its mean, and increasing the latter, which has nothing to do with goodness-of-fit, can also increase R^2. For example, the explanatory variables may be chosen such that half of them are in one closely spaced group and the other half in another closely spaced group, with the two groups spaced widely apart. This disposition of the Xs tends to make the denominator of R^2 large, while having no effect whatsoever upon how well the regression model fits the observed data.

For linear regression models *without* an intercept, such as

$$Y = \beta X \tag{4.9}$$

or

$$Y = \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_p X_p \tag{4.10}$$

R^2 cannot function as a goodness-of-fit criterion without modification. The modification is usually made by defining R^2 to be the complement of the ratio of the sum of squares of the residuals to the *uncorrected* sum of squares, where the least-squares regression line is determined by forcing the line to pass through the origin, i.e., the point (0,0) on the (X,Y) axis.

Attempts have been made to generalize R^2 by correcting it in such a way that it becomes appropriate for nonlinear regression as well as for models without an intercept, including models with stochastic assumptions other than the normal (Nagelkerke, 1991). Rather than looking at the question from a theoretical point of view, which involves complicated mathematics, we will look at a practical example. Consider the data set for CLD38 using the square-root model with the stochastic assumption that the variance is homogeneous in $\sqrt{\mu}$. From Table 4.4, the residual mean square was 5.924×10^{-6} for this data set. Carrying out the least-squares regression analysis in the usual way, and presenting the results in the form of an analysis of variance (ANOVA) table, leads to the following tabulation.

Source of Variation	df	Sum of Squares	Mean Squares	Approx. *F*	*Pr* > *F*
Regression	4	0.7458	0.1864	31474	<0.0001
Residual	14	0.000083	5.924E-6		
Corrected total	17	0.0826			
Uncorrected total	18	0.7459			

If one were to define the coefficient of determination in the usual way as the complement of the residual sum of squares divided by the sum of squares corrected for the mean, one would obtain $R^2 = 1 - 0.000083/0.0826 = 0.999$. This would suggest that 99.9% of the "explainable" variation in the response variable $\sqrt{\mu}$ has been accounted for by the square-root nonlinear regression model, a suspiciously high figure. The inappropriateness of R^2 for nonlinear regression models, whether they have an intercept or not, has been dealt with previously (Ratkowsky, 1983, 1990). In the above calculation, the corrected total sum of squares is the sum of squares of the response variable adjusted for its mean. Although centering around the mean is justifiable when the model is a straight line or a plane that passes through the mean, there is no heuristic reason why it should be a sensible procedure when the model is nonlinear. Despite warnings of its potentially misleading nature, R^2 continues to be misused not only in the predictive microbiology literature but also wherever nonlinear regression models are employed.

Another widely used criterion is the "percentage variance accounted for" or "adjusted R^2," which differs from the traditional R^2 in being based upon the mean squares rather than the sum of squares; that is, this alternative measure is defined as the complement of the ratio of the residual mean square to a mean square based upon the corrected total. In the above example, this adjusted R^2 would be

$$\text{Adjusted } R^2 = 1 - 0.000005924/(0.0826/17) = 0.9988$$

a closely similar result to that for the traditional R^2. This too is inappropriate for a nonlinear regression model. This criterion was used by Davey and Amos (2002), but its use was appropriate, because there the model in question was a linear regression model with an intercept.

Another widely used but inappropriate indicator of goodness-of-fit both for linear and nonlinear regression models that often appears in the predictive microbiology literature is a plot of predicted response vs. observed response. Although it might appear that predicted responses are "close" superficially to the 45° line, there may also be a clear pattern of discrepancy manifested by long runs of like-signed residuals, which are the differences between the observed and the predicted values. Misleading inferences may easily be made by the indiscriminate use of such graphs.

4.2.5.2 Root Mean Square Error

Probably the most simple and the most informative measure of goodness-of-fit for regression models, both linear and nonlinear, is the root mean square error (RMSE), defined as the square root of the residual mean square. The RMSE may be viewed as the "average" discrepancy between the observed data, transformed if necessary, and their predicted values. Hence, its magnitude, especially when one also considers

the precision in the original data, is useful in assessing whether a given model truly fits the data well. For the data sets from Table A4.1 and Table A4.2, the RMSEs are simply the square roots of the entries in Table 4.2 and Table 4.4, respectively.

The most desirable situation occurs when each experimental condition in the whole experiment is replicated, as that will enable one to calculate a measure of precision from the contributions of the replications to the residual variance. If the variances are of similar magnitude for each experimental condition, a pooled variance may be calculated. If the magnitudes of the residual mean square and the pooled variance are similar, this suggests that the model fits the data well. If the residual variance is much larger than the pooled variance, improvements to the model should be sought. When experiments are not replicated, the data required to calculate the pooled variance are not available. For the data given in Table A4.1, a pair of data sets is available, and if it can be shown that the lag times obtained from the breast data are not significantly different from those obtained from the thigh data, the two data sets may be pooled. We will now formally carry out the tests of significance to test the null hypothesis that the breast and thigh data sets are closely similar.

Since the extended hyperbola model, coupled with the log rate stochastic assumption as model (4.2a), seemed to be best (from the residual mean squares of Table 4.2), we will use that model to illustrate the procedure for testing whether the breast and thigh data may be pooled. Fitting 4.2a to the data sets separately produces regression results of which the following ANOVA table may be extracted:

Source of Variation	df	Sum of Squares	Mean Squares
Breast Meat Data (n = 21)			
Due to regression	3	49.8649	16.6216
Residual	18	0.4937	0.0274
Corrected total	20	23.3117	
Uncorrected total	21	50.3585	
Thigh Meat Data (n = 21)			
Due to regression	3	46.8351	15.6117
Residual	18	0.1696	0.0094
Corrected total	20	21.8457	
Uncorrected total	21	47.0047	

It is clear from the high ratio of regression to residual mean squares that the model fits the data well for each data set. The two sets of data are now combined into a single pooled data set of size $n = 42$, and model 4.2a fitted to the combined set. The following ANOVA table extract is obtained.

Source of Variation	df	Sum of Squares	Mean Squares
Combined Data (n = 42)			
Due to regression	3	96.6415	32.2138
Residual	39	0.7217	0.0185
Corrected total	41	45.1745	
Uncorrected total	42	97.3633	

We can now compare the residual sum of squares of 0.7217 with 39 df to the "pooled" residual sum of squares of 0.4937 + 0.1696 = 0.6633 with 18 + 18 = 36 df. The difference between these two sums of squares is 0.7217 – 0.6633 = 0.0584 with 39 – 36 = 3 df. This leads to the following variance ratio test:

$$F_{3,36} = \frac{0.0584/3}{0.6633/36} = 1.057$$

This variance ratio of 1.057 has an F distribution with 3 and 36 df and is clearly nonsignificant; hence, the breast and thigh data sets may be combined. We note that the residual mean square for the combined data, 0.0185, is almost identical to the pooled residual mean square of 0.0184, so that we have no hesitation in pooling these two sets of data into a single combined set.

Even if the entire experiment cannot be replicated, there is merit in trying to replicate some of the experimental conditions in one's experiment. Doing so provides one with a pooled error against which the residual mean square may be formally tested using the variance ratio test.

4.2.5.3 Examination of Residuals

Examination of the residuals is an important component of the evaluation of regression models, enabling the user to assess whether the model fits the data adequately. A residual is defined as the difference between the observation and the fitted or predicted value,

$$r_i = y_i - \hat{y}_i$$

where r_i is the residual corresponding to the ith observation y_i, and $\hat{y}_i$ is the corresponding predicted value. Commonly used techniques for examining residuals include plots of residuals vs. predicted values, normal probability plots, and calculating measures of influence. These procedures are described in books such as those by Mendenhall and Sincich (1996) and Fox (1991), and are carried out by software packages such as SAS (1990). Employing plots of residuals vs. predicted values and normal probability plots and associated tests enables the modeler to examine the assumptions inherent in regression analysis, such as normality of the residuals and equality of the error variance. In particular, they readily identify outlying observations, some of which may be data entry errors. Measures of influence extend the examination further, shedding further light on unusual observations.

4.2.5.3.1 The Runs Test

A simple first step in the examination of residuals is to order the residuals so that they are arranged according to increasing order of the explanatory variable X (also referred to as the "independent" or "regressor" variable), and then count the number of runs of like-signed residuals. The more runs there are, the more the fitted model tends to be centrally located within the set of data points, and thus the better the

TABLE 4.5
Number of Runs of Like-Signed Residuals[a]

Model	Assumption		
	Untransformed	Log Rate	Square-Root Rate
(4.1) Hyperbola	9	7	7
(4.2) Extended hyperbola	20	20	20
(4.3) Linear Arrhenius (*T* in Kelvin)	8	20	16
(4.4) Simple square root	10	13	15

[a] Data of Table A4.1.

goodness-of-fit. The runs test was used by Oscar (2002), who failed to mention that an ambiguity arises when there is more than one observation at an *X* value. For the data in Table A4.1, there are two observations at each of the 21 temperatures, if, as appears justified for these data, the data for breast lag time are pooled with the data for thigh lag time. For want of a better procedure, we arrange the data in such a way so that we start with the breast measurement at 8°C, follow it with the thigh measurement at 8°C, and continue alternating breast and thigh measurements until the thigh measurement at 48°C is reached.

Results of applying the runs test to the fitted models for the data of Table A4.1 are presented in Table 4.5. The model with a consistently poor fit is the "hyperbola" model (4.1), which has few (range 7 to 9) runs irrespective of the stochastic assumption, and that with a consistently good fit is the "extended hyperbola" model (4.2), with 20 runs of like-signed residuals for each error assumption. The figures for the "Linear Arrhenius" model (4.3) are interesting. When the response variable is untransformed, so that lag time λ is modeled directly, the fit is poor (8 runs), but when log rate is used, as intended by Davey (1989, 1991), the fit is excellent (20 runs). It is less good (16 runs) when the square-root transformation is applied to the rate instead of log rate. The simple square-root model (4.4) occupies an intermediate position. Although at its best when used with the square-root transformation of the rate, its 15 runs of like-signed residuals do not compete with models 4.2 or 4.3 when the log rate assumption is used.

Further results of the examination of the residuals are presented in Table 4.6. The tabulated *P* values confirm that the residuals are severely nonnormally distributed for all models when the response variable is untransformed and the lag time λ is modeled directly. The reason for this is that when untransformed, the large lag times that occur at temperatures of 8, 10, 12, and 14°C are not only poorly fitted by all four models, but also result in residuals that are often an order of magnitude larger than those at the higher end of the temperature scale. The results show that model 4.3 produces normally distributed residuals with the log rate transformation and the square-root transformation. What is surprising is the nonnormal distribution of the residuals for model 4.2 with the log rate stochastic assumption, which was the clear winner in terms of goodness-of-fit, well ahead of the other three models with that assumption. This result is due to a single data point, the lag time value of

TABLE 4.6
Probability Values Associated with the Test of Normality of the Residuals[a]

Model	Assumption		
	Untransformed	Log Rate	Square-Root Rate
(4.1) Hyperbola	<0.0001	0.489	0.020
(4.2) Extended hyperbola	<0.0001	0.0002	<0.0001
(4.3) Linear Arrhenius (*T* in Kelvin)	<0.0001	0.697	0.268
(4.4) Simple square root	<0.0001	0.010	<0.0001

[a] Data of Table A4.1.

TABLE 4.7
Probability Values Associated with the Test of Normality of the Residuals[a]

Model	Assumption		
	Untransformed	Log Rate	Square-Root Rate
(4.1) Hyperbola	<0.0001	0.450	0.027
(4.2) Extended hyperbola	<0.0001	0.128	0.031
(4.3) Linear Arrhenius (*T* in Kelvin)	<0.0001	0.471	0.357
(4.4) Simple square root	<0.0001	0.858	0.240

[a] Data of Table A4.1, with the data point for breast meat at 48°C eliminated.

1.6 h on breast meat at 48°C. Compared with all other values of lag time for temperatures above 35°C, that value is clearly too high (see Table A4.1). Eliminating that data point and refitting the model results in a set of normally distributed residuals, not only for model 4.2 but also for model 4.4, as shown in Table 4.7. Similarly, the table shows that use of model 4.4 in combination with the square-root stochastic assumption results in a set of normally distributed residuals.

4.2.5.3.2 Measures of Influence

One of the statistics measuring influence is the so-called "hat matrix" H, identifying those observations that are influential due to the values of the explanatory variable(s). An analogy is to consider children sitting on a seesaw. The further they sit from the fulcrum, the greater the "leverage," the word often employed to describe h_i, the ith element of the diagonal of that matrix. Since $\hat{y}_i$ may be written as a linear combination of the n observed y values,

$$\hat{y}_i = h_1 y_1 + h_2 y_2 + ... + h_n y_n$$

the larger the value of h_i, the greater the weight given to the ith observation. The average value of h is conveniently given by the ratio of the number of parameters in the model to the total number of points n.

Another useful statistic is the Studentized residual, which is the ratio of the ordinary residual r_i to its standard error, the latter incorporating the leverage measure h_i. High values of this statistic, greater than two (say) in absolute magnitude, indicate a significantly large residual. Other measures of influence include Cook's D (Cook, 1979) and the Dffits statistic (Belsley et al., 1980). These statistics produce a combined measure of influence by coupling the effect of high leverage with the measure of whether the observation is an outlier. Hence, a large value of D or Dffits usually results from both the leverage and the residual being large. D_i is customarily compared to critical values of the F distribution with numerator df equal to the number of parameters estimated and denominator df equal to the residual df. If D_i exceeds the 50th percentile of this F distribution, the observation is deemed to be influential (see Mendenhall and Sincich, 1996).

Measures such as Cook's D and Dffits are intended for use with the "straight-line" model (4.7) or with the multiple regression model (4.8), just in the same way that R^2 or adjusted R^2 are intended to assess goodness-of-fit for such models. We have seen in Section 4.2.5.1 that moving the "fulcrum" from the center of the line or the plane to the origin of the coordinates results in an incorrect interpretation of R^2 or adjusted R^2 if the standard definition of those measures is not modified. Similarly, influence has to do with the distance that a point is from the fulcrum, and whereas such a distance is unambiguous with models such as 4.7 and 4.8, various problems of interpretation arise when one is dealing with a curvilinear regression, such as the polynomial models to be discussed in Section 4.3.1 or nonlinear regression models, ones like 4.5 and 4.6, in which the parameters appear nonlinearly. In any event, many of the methods for examining residuals are graphically based (such as normal probability plots and graphs of residuals vs. fitted values), and tests of significance should be considered to be only approximate. This is especially true for nonlinear regression models because of bias in the predicted response values, although such bias is typically small (see Ratkowsky, 1983, for discussion of the effect of the so-called "intrinsic" nonlinearity).

Table 4.8 shows some results of applying measures of influence to the data sets of Table A4.2, using the square-root model (4.5) coupled with the square-root stochastic assumption. Similar to the results shown in Table 4.3 and Table 4.4, little difference was observed between the parameter estimates obtained from the three stochastic assumptions, as well as between models 4.5 and 4.6.

For the CLD38 data set, the average value of the leverage h_i is 4/18 = 0.222, since there are four parameters in the model and a total of 18 data points. Since 2(0.222) = 0.444, only the last data point, the one corresponding to t = 29.9°C, exceeds this value and appears to be influential. Although the Studentized residual is far below 2.0, Cook's D of 2.43 exceeds the critical value of 1.52 for the F distribution with 4 and 14 df for the 50th percentile. Thus, the last data point should be considered to be significantly influential without being an outlier. There are indications that two of the interior points, the sixth and the ninth observations, have high residuals. This is confirmed by looking at Figure 4.2a, which shows a very

TABLE 4.8
Results of the Examination of Residuals from Use of Model 4.5 with the Square-Root Stochastic Assumption for the Data of Table A4.2[a]

Obs	Temp	Square Root of Rate	Predicted Square Root of Rate	Residual	Leverage h_i	Studentized Residual	Cook's D
				CLD 38			
1	1.3	0.07727	0.07535	0.00192	0.333	0.965	0.1164
2	2.8	0.08972	0.09030	–0.00058	0.229	–0.272	0.0055
3	4.0	0.10178	0.10224	–0.00046	0.170	–0.207	0.0022
4	5.2	0.11265	0.11415	–0.00150	0.131	–0.662	0.0165
5	7.2	0.13198	0.13391	–0.00192	0.104	–0.835	0.0202
6	7.9	0.14632	0.14079	0.00553	0.104	<u>2.402</u>	0.1671
7	9.2	0.15215	0.15349	–0.00134	0.113	–0.586	0.0109
8	11.4	0.17461	0.17472	–0.00011	0.145	–0.049	0.0001
9	12.4	0.17961	0.18422	–0.00461	0.161	–<u>2.066</u>	0.2047
10	17.2	0.22989	0.22729	0.00260	0.176	1.178	0.0740
11	18.3	0.23782	0.23621	0.00161	0.163	0.723	0.0254
12	19.5	0.24413	0.24531	–0.00117	0.148	–0.523	0.0119
13	21.7	0.25994	0.25957	0.00037	0.148	0.164	0.0012
14	23.0	0.26556	0.26598	–0.00043	0.179	–0.193	0.0020
15	24.3	0.27137	0.27031	0.00106	0.232	0.498	0.0188
16	25.6	0.27216	0.27191	0.00025	0.287	0.122	0.0015
17	28.1	0.26171	0.26406	–0.00235	0.315	–1.168	0.1566
18	29.9	0.24648	0.24534	0.00114	<u>0.861</u>	1.253	<u>2.4251</u>
				16L16			
1	0.0	0.07668	0.07469	0.00199	0.231	0.769	0.0444
2	1.5	0.09241	0.09078	0.00163	0.184	0.611	0.0210
3	2.4	0.10407	0.10043	0.00363	0.160	1.340	0.0852
4	4.1	0.11730	0.11867	–0.00137	0.122	–0.494	0.0085
5	5.2	0.12783	0.13047	–0.00264	0.103	–0.943	0.0256
6	7.1	0.14761	0.15085	–0.00323	0.081	–1.140	0.0287
7	7.9	0.16331	0.15943	0.00388	0.075	1.366	0.0381
8	9.3	0.17018	0.17443	–0.00426	0.071	–1.493	0.0426
9	11.3	0.19519	0.19586	–0.00067	0.075	–0.234	0.0011
10	13.3	0.21394	0.21725	–0.00331	0.089	–1.172	0.0337
11	14.3	0.22709	0.22792	–0.00083	0.099	–0.295	0.0024
12	17.4	0.26315	0.26080	0.00236	0.129	0.853	0.0270
13	18.5	0.27421	0.27233	0.00187	0.136	0.682	0.0182
14	19.5	0.27842	0.28272	–0.00429	0.138	–1.563	0.0978
15	22.0	0.31190	0.30791	0.00398	0.130	1.444	0.0782
16	23.1	0.31859	0.31843	0.00016	0.129	0.059	0.0001
17	24.5	0.33332	0.33093	0.00238	0.145	0.872	0.0322
18	25.6	0.33985	0.33973	0.00012	0.188	0.046	0.0001
19	28.1	0.35440	0.35372	0.00068	<u>0.381</u>	0.293	0.0133
20	29.9	0.35071	0.35427	–0.00356	<u>0.405</u>	–1.560	0.4142

TABLE 4.8 (Continued)
Results of the Examination of Residuals from Use of Model 4.5 with the Square-Root Stochastic Assumption for the Data of Table A4.2[a]

Obs	Temp	Square Root of Rate	Predicted Square Root of Rate	Residual	Leverage h_i	Studentized Residual	Cook's D
21	31.6	0.34220	0.34075	0.00145	<u>0.928</u>	1.821	<u>10.6762</u>
				M68			
1	19.0	0.04508	0.04554	−0.00046	0.327	−0.520	0.0329
2	19.8	0.04817	0.04736	0.00081	0.270	0.875	0.0708
3	21.5	0.05158	0.05121	0.00036	0.176	0.371	0.0074
4	23.5	0.05407	0.05573	−0.00166	0.111	−1.623	0.0826
5	24.7	0.05812	0.05843	−0.00031	0.094	−0.303	0.0024
6	25.6	0.06131	0.06045	0.00086	0.091	0.833	0.0173
7	26.8	0.06325	0.06312	0.00013	0.096	0.123	0.0004
8	28.2	0.06537	0.06619	−0.00082	0.112	−0.805	0.0204
9	29.3	0.06868	0.06855	0.00013	0.127	0.126	0.0006
10	30.3	0.07198	0.07064	0.00134	0.140	1.332	0.0721
11	31.6	0.07392	0.07323	0.00069	0.149	0.692	0.0210
12	33.0	0.07647	0.07576	0.00072	0.148	0.716	0.0223
13	34.5	0.07692	0.07797	−0.00105	0.143	−1.046	0.0457
14	36.0	0.07809	0.07931	−0.00122	0.166	−1.237	0.0763
15	37.4	0.07809	0.07923	−0.00114	0.242	−1.206	0.1164
16	38.8	0.07906	0.07697	0.00209	0.337	<u>2.365</u>	0.7107
17	40.0	0.07198	0.07224	−0.00026	0.344	−0.300	0.0119
18	41.5	0.06019	0.06039	−0.00020	<u>0.925</u>	−0.660	1.3528

[a] Significant statistics are underlined.

good fit of the square-root model (and the cardinal temperature model as well), with these two points being further from the fitted curve than any others.

For the 16L16 data set, the average value of h_i is 4/21 = 0.190, so the last three data points, especially the last one, have significant leverage. This is reflected in the highly significant value of Cook's D for the last point, but there do not appear to be any outliers. Figure 4.2b confirms this by displaying a very close fit between the square-root model and the data.

For the M68 data set, the average value of h_i is 0.222, so that the last data point exerts considerable leverage. Nevertheless, Cook's D of 1.35 is below the critical value of 1.52, so this point is not unduly influential. Observation 16 is seen to have a Studentized residual in excess of 2.0, which is confirmed by looking at Figure 4.2c.

We mention once again that the examination of residuals is an important tool, but that one should rely on graphical interpretation more than on significance testing, since the models are nonlinear regression models. Measures such as Cook's D are not strictly applicable, and like R^2 or adjusted R^2, they may be inappropriate or misleading for nonlinear regression models.

4.2.5.4 Measures of Nonlinear Behavior

Fitting of nonlinear regression models has become a relatively straightforward task with the use of modern statistical packages. However, regression modeling should not be viewed simply as a curve-fitting exercise, but one that requires thought and subsequent evaluation and testing. The examination of residuals, for example, which was the subject of the previous section, is part of the process of evaluation that logically follows the routine fitting of a mathematical model to a data set. A further step in that process is to ask whether there are other features of the model that may or may not be deemed desirable in a mathematical model. When the fitted model is a nonlinear regression model, one should ask whether the model is "close to linear" or not.

The concept of a "close to linear" nonlinear regression model was advanced in an earlier book (Ratkowsky, 1983). It was recognized then that some nonlinear regression models could have severely biased parameter estimates, have a probability distribution that was vastly different from that of a normal (Gaussian) distribution, typically being skewed with a long right-hand or left-hand tail, and have excess variance. This contrasts with linear regression models such as 4.7 and 4.8, which, when the stochastic assumption of a normally distributed error term is valid, have unbiased, normally distributed, minimum variance estimators. Although all nonlinear regression models have biased parameter estimators, the various models differ in the extent of the bias. The models that have only a very small bias in their estimates were called "close to linear" by Ratkowsky (1983), whereas those that exhibited severe bias were said to be "far from linear." In many models, parameter bias may be reduced by reparameterization, i.e., changing the form in which the parameters of the models appear. Other models can only be reparameterized at the price of producing an awkward-appearing model. (See Ratkowsky [1983, 1990] for a detailed discussion of these issues and for many examples of reparameterization.)

Several measures of nonlinear behavior have been advanced over the years, some of which have not withstood the test of time. A very reliable indicator of nonlinear behavior for an individual parameter estimator is based on Hougaard's measure of skewness, described in Ratkowsky (1990, pp. 27–28), which exploits the close connection between a nonlinear regression model's behavior and its expression in biased, skewed parameter estimators. This measure is available in recent releases of SAS© statistical software, PROC NLIN, using the option "Hougaard." Experience with this measure suggests that if the Hougaard skewness measure is less than 0.1 in absolute value, the estimator of the parameter is very close to linear, but that if its absolute value exceeds 0.25, the skewness is quite apparent (as may be seen, for example, by carrying out a simulation study), and if it exceeds 1.0, considerable nonlinear behavior of the estimator is present. Since skewness and bias (the difference between the mean value of a parameter's estimator and its true population value) are closely correlated, a high skewness measure can be taken to mean a high bias in the estimator of that parameter, and conversely, a low skewness measure equates to a low bias.

Table 4.9 presents results for Hougaard's skewness measure for the parameters of models 4.1 to 4.4, in combination with the data of Table A4.1 for the various

TABLE 4.9
Hougaard Skewness Measures for Models 4.1 to 4.4 in Their Original and Transformed Forms, Pooled Data of Table A4.1 for the Different Stochastic Assumptions

Model	Parameter	Assumption		
		Untransformed	Log Rate	Square-Root Rate
(4.1) Hyperbola	p	0.063	0.207	0.219
	q	–0.077	–0.360	–0.096
(4.2) Extended hyperbola	p	0.384	0.216	0.207
	q	–0.390	–0.540	–1.178
	m	0.346	0.279	0.632
(4.3) Linear Arrhenius	A	0.169	0	–0.055
	B	–0.192	0	0.053
	C	0.218	0	–0.051
(4.4) Simple square-root	b	0.119	0.017	0
	T_{min}	–0.089	–0.217	–0.214

stochastic assumptions. Although the extended hyperbola model (4.2) containing an exponent m fits the data much better than the simple hyperbola model (4.1), the parameter estimators for p and q are much more biased than they were for 4.1. For example, q substantially underestimates that parameter. The Linear Arrhenius model (4.3) has zero bias when the log assumption is used, reflecting the fact that that model is a linear regression model, and it has a small, nonperceptible bias for the square-root stochastic assumption. The simple square-root model also has parameters with low bias (the zero value for b with the square-root stochastic assumption reflecting the fact that the model is linear).

Table 4.10 presents results for Hougaard's skewness measure for the parameters of models 4.5 and 4.6, in combination with the three sets of data of Table A4.2 for the various stochastic assumptions. The four-parameter square-root model (4.5) and the cardinal temperature model (4.6) both contain the notional parameters T_{min} and T_{max} and the results for both models are in agreement in that the biases in T_{min} are both small, whereas the biases for T_{max} are quite large for all three data sets, particularly so for the 16L16 data. The bias for both models is positive, meaning that the estimates, on average, are larger than the true values.

4.2.6 Building Mathematical Models

This section takes a look at the construction of models for predictive microbiology. Over the years, a variety of opinions have been expressed about the nature of models that might be used to describe the shelf life of food products and the rate at which food deteriorates. Many of these opinions are philosophical in nature. For example, some authors have been concerned with questions such as the differences between mechanistic and empirical models, among others (e.g., see Section 4.2.6.3). Herein, we will confine our attention to considerations that have led to the appearance of

TABLE 4.10
Hougaard Skewness Measures for Models 4.5 and 4.6 Fitted to the Data of Table A4.2 for the Different Stochastic Assumptions

Model	Parameter	Assumption		
		Square-Root Rate	Rate	Log Rate
CLD38 (Temperature Estimates in Kelvin)				
(4.5) 4-Parameter square root	T_{min}	0.001	0.002	0.019
	T_{max}	0.281	0.241	0.388
	b	0.274	0.284	0.284
	c	0.190	0.134	0.280
(4.6) Cardinal temperature	T_{min}	–0.038	–0.027	–0.026
	T_{max}	0.429	0.343	0.606
	T_{opt}	0.003	–0.006	0.021
	μ_{opt}	0.062	0.029	0.126
16L16 (Temperature Estimates in Kelvin)				
(4.5) 4-Parameter square root	T_{min}	0.012	0.030	0.023
	T_{max}	0.421	0.332	0.668
	b	0.197	0.219	0.195
	c	0.326	0.207	0.670
(4.6) Cardinal temperature	T_{min}	–0.006	0.007	0.024
	T_{max}	0.742	0.562	1.121
	T_{opt}	0.072	0.060	0.032
	μ_{opt}	0.137	0.069	0.314
M68 (Temperature Estimates in Kelvin)				
(4.5) 4-Parameter square root	T_{min}	–0.091	–0.098	–0.078
	T_{max}	0.342	0.363	0.330
	b	0.194	0.239	0.176
	c	0.221	0.213	0.245
(4.6) Cardinal temperature	T_{min}	–0.133	–0.138	–0.122
	T_{max}	0.434	0.451	0.431
	T_{opt}	0.014	0.006	0.025
	μ_{opt}	0.020	0.003	0.044

various classes of empirical models for practical use that have appeared in the predictive microbiology literature.

4.2.6.1 Why Polynomial Models Do Not Work

One class of models that is frequently encountered in the predictive food microbiology literature is "polynomial models." For example, if one were modeling specific growth rate constant (μ) as a function of temperature, modelers favoring polynomial models would use, instead of models such as 4.5 and 4.6, a model in which μ is expressed as a low-order polynomial in T, usually not exceeding the third order, i.e.,

$$\mu = a + bT + cT^2 + dT^3 \tag{4.11}$$

When there is more than a single environmental factor, the number of parameters of such models multiplies rapidly. For example, if temperature and salt concentration (NaCl) are the environmental factors, the model would become, if all terms of the polynomial up to third order are included,

$$\begin{aligned}\mu = a + bT + cT^2 + dT^3 + e\,(\text{NaCl}) + f(\text{NaCl})^2 + g(\text{NaCl})^3 \\ + hT(\text{NaCl}) + iT^2(\text{NaCl}) + jT^3(\text{NaCl}) + lT(\text{NaCl})^2 + mT^2(\text{NaCl})^2 \\ + nT^3(\text{NaCl})^2 + oT(\text{NaCl})^3 + pT^2(\text{NaCl})^3 + qT^3(\text{NaCl})^3\end{aligned} \tag{4.12}$$

A total of 16 parameters (coefficients) have to be evaluated for this complete third-order polynomial. This lack of parsimony often compels authors, for practical reasons, not to go beyond second-order terms, so that the model becomes

$$\begin{aligned}\mu = a + bT + cT^2 + d(\text{NaCl}) + e(\text{NaCl})^2 + fT(\text{NaCl}) + gT^2(\text{NaCl}) \\ + hT(\text{NaCl})^2 + iT^2(\text{NaCl})^2\end{aligned} \tag{4.13}$$

which has "only" nine parameters. Aside from its nonaesthetic appearance, there remains the practical question of whether such a model is capable of fitting real data.

As an example of the nonparsimonious nature of polynomials, let us look at one of the illustrative data sets of Table A4.1, that for CLD38. The graphs in Figure 4.3 display the fit of a quadratic, a cubic, and a quartic (i.e., fourth degree) polynomial, using the "rate" assumption (i.e., no transformation), and also show the observed data points. The quadratic fit, involving three parameters, is very poor, there being only four runs of like-signed residuals, with the predicted model underestimating the observed data at very low temperatures and at temperatures near the optimum, and

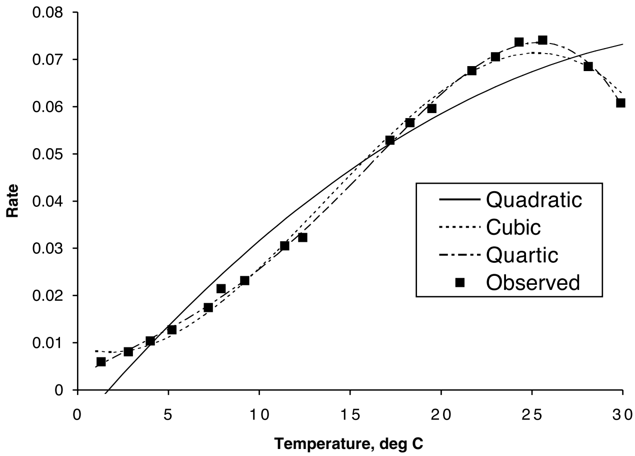


FIGURE 4.3 Observed and predicted rates (fitted quadratic, cubic, and quartic polynomials) vs. temperature. CLD38 data from Table A4.2.

overestimating at moderate temperatures and very high temperatures. The cubic fit, with four parameters, is a considerable improvement, but there are still only five runs of like-signed residuals, and the residual mean square is 3.67×10^{-6}, much higher than those for the square-root or cardinal temperature models. The quartic fit, with five parameters, does very well with 11 runs of like-signed residuals, with a residual mean square of 0.767×10^{-6}, which is almost as low as that of the square-root model and lower than that of the cardinal temperature model (see Table 4.4). To achieve this precision, however, one extra parameter beyond that required by the square-root or cardinal temperature models had to be employed, and none of the five parameters of the quartic polynomial is interpretable. In contrast, the square-root model has two interpretable parameters and the cardinal temperature model has four.

4.2.6.2 Comparison of Models Using *F* Tests

A frequently used test in the statistical literature for comparing models is the *F* test, employed extensively for formal testing of ANOVA models. In the food microbiology literature, it has been used, for example, to describe the combined effects of temperature, pH, and lactate on the growth of *Listeria innocua* (Houtsma et al., 1996), to quantify the interactions of spoilage microorganisms (Pin and Baranyi, 1998) and to determine if a simple, nested model was sufficient to describe the growth kinetics of a number of microorganisms (Delignette-Muller, 1998). Some limitations of this method have been noted (McMeekin et al., 1993), for example, (1) it cannot discriminate between models with the same number of parameters, or nonnested models; (2) the significance of the *F* test is only approximate for nonlinear regression models; (3) indiscriminate use of the *F* test may lead to overparameterized models, i.e., ones with more terms and parameters than are necessary. Some of these limitations are more serious than others; for example, the approximate nature of the *F* test in nonlinear regression models is not serious, as the bias in the *F* test depends upon the component of nonlinearity referred to as the "intrinsic" nonlinearity, and this bias is typically small in most nonlinear regression models, except for very small sample sizes (see Ratkowsky, 1983).

Models that typically are overparameterized are the polynomial models, criticized in Section 4.2.6.1. Some authors have tried to reduce the number of parameters by eliminating nonsignificant terms (e.g., Houtsma et al., 1996), but it is not clear what purpose is really served by that procedure. First, for correctness, the eliminated terms must be *jointly* nonsignificant, a conclusion that cannot be reached by applying stepwise procedures such as forward or backward elimination. One should compare the reduced model with the full model, which may be done using the *F* test, to see whether inclusion of the extra terms significantly improves the fit. In any event, the final model after terms have been eliminated is really no "better" than it was before the unnecessary terms were deleted; that is, although there may appear to be less unexplained variation in the response variable due to a smaller residual mean square, because there are now more df for error than before, the effect is mainly cosmetic. There is no substitute, when the goal is to produce accurate rate models or growth/no growth interface models for predicting food product shelf life, for trying to build the best possible mechanistic model for the process, or a close approximation to it.

4.2.6.3 Models with Several Environmental Factors

Although the earliest successful models in food microbiology involved only temperature, including the Bĕlehrádek-type models of which the square-root model is a special case (Ratkowsky et al., 1982, 1983) and the Arrhenius-type relationships (Schoolfield et al., 1981; Sharpe and DeMichele, 1977), it soon became apparent that other growth-limiting environmental factors had to be taken into account. Models containing water activity (a_w) in addition to temperature followed (e.g., Davey, 1989; McMeekin et al., 1987), and with time, the effect of hydrogen ion concentration in the form of pH (e.g., Adams et al., 1991) and the addition of weak acids such as acetic acid and lactic acid were being considered (e.g., Presser et al., 1997).

A parsimonious model involving the combined effect of temperature, water activity (or salt concentration), and the addition of a weak acid, such as lactic acid, cannot be successfully achieved using polynomials such as 4.11 to 4.13. Baranyi et al. (1996) promoted the desirability of models embodying known or assumed features of the phenomenon under consideration. Van Impe et al. (2001) considered models to be divided into three classes, following Ljung (1999), as white box models, black box models, and gray box models. Deductive white box models require full knowledge of the underlying physical mechanisms and a deep understanding of the physical, chemical, and biochemical laws driving the process, a situation that is rarely available at this moment in time. Black box models lie at the opposite end of the scale. They take the experimental data as input information and produce output variables with or without necessarily producing an equation or series of equations. This inductive approach includes polynomial modeling and the use of artificial neural networks, but models so produced cannot reflect physical considerations. A gray box model is a compromise between the two extremes and is probably the standard to which modelers in predictive food microbiology can realistically hope to achieve at this point of time. Another alternative, suggested by Geeraerd et al. (2002), is to retain the black box approach while incorporating *a priori* microbiological knowledge into the modeling process so that overfitting of the data and unrealistic parameter estimates are prevented from occurring.

The approach taken by Presser et al. (1997) was an attempt to incorporate some reasonable assumptions based upon physical chemistry into the modeling process. They used the observation reported by Cole et al. (1990) that the growth rate of a microorganism is directly proportional to the hydrogen ion concentration, and this led directly to an expression for the effect of pH. Similarly, the well-known Henderson–Hasselbalch equation of physical chemistry was used, which related the ratio of the undissociated to the dissociated forms of a weak organic weak to the pH and pK_a, the latter being the pH at which the concentrations of the two forms are equal. The resulting growth rate model (see Presser et al., 1997) for *E. coli* as a function of temperature, pH, and added lactic acid concentration contained only six parameters to be estimated. This is in sharp contrast to a polynomial model, which would have had to contain dozens of parameters to achieve the same level of prediction. If the model fit exhibits shortcomings, then it can be amended to improve its predictive ability, but the basic model form is a good foundation upon which to base further fine-tuning.

4.2.7 Replicated Data Sets

We now examine the data in Table A4.3, which, unlike the data in Table A4.1, can be seen to be a group of data sets with genuine replication. The data of Table A4.1 served as a surrogate for replication, since the results for breast meat were indistinguishable from those on thigh meat, making it possible to consider the two sets of data to be replicates. The growth rate data for *L. monocytogenes* of Table A4.3 were obtained from five separate runs using a temperature gradient incubator, on samples of what was ostensibly the same material. Each data set is independent of the others.

The four-parameter square-root model (4.5) appears to be well suited to fit each of the individual data sets, as may be seen from Figure 4.4. Parameter estimates and their standard errors are given in Table 4.11. It is seen that there is a fair amount of variability in the estimates of the two cardinal temperatures, with the T_{min} estimates varying between –1.0 and 2.3°C and the T_{max} estimates varying between 45.5 and 48.3°C. Similarly, the measures of goodness-of-fit show considerable variation, with the residual mean squares ranging by a factor of 3 between 1.0×10^{-5} and 3.0×10^{-5} and the number of runs of like-signed residuals varying between only six runs for Data Set 4 to a rather substantial 20 runs for Data Set 3. Nevertheless, there is no correlation between number of runs of residuals and the normality of the set of residuals, with Data Set 4 having a close-to-normal set of residuals and Data Sets 2 and 3 being marginally nonnormal. As can be seen from Figure 4.5, which shows the pooled Data Sets 1 to 5 on a single graph with the square-root model (4.5) fitted to the pooled data, there is a group of five data points in Data Set 1 in the suboptimal temperature range of 27.5 to 32.8°C with much lower rates than those predicted by the overall fitted model.

Pooling the residual sum of squares from the individual data sets leads to 0.000661 + 0.000493 + 0.000333 + 0.000478 + 0.000250 = 0.002215, with 22 + 23 + 23 + 23 + 24 = 115 df. The residual sum of squares for the pooled data set of 135 points is 0.00677 (see Table 4.11) with 131 df. The difference between these two sums of squares is 0.00677 – 0.002215 = 0.00455 with 131 – 115 = 16 df. This leads to the following variance ratio test:

$$F_{16,115} = \frac{0.00455/16}{0.002215/115} = 14.76.$$

This is clearly a highly significant F value ($P < 0.001$) and indicates model inadequacy. The question is whether this is due to a poor model or poor data. The information in Figure 4.5 suggests that the model is adequate but that there are a number of aberrant data points. This is further borne out by Figure 4.6, which is a plot of the residuals vs. the fitted values from fitting 4.5 to the pooled set of 135 data points. There are seven data points that stand out as potential outliers, with residuals exceeding 0.015 in absolute magnitude. These are indicated using a larger font size.

Removing the data points with the seven biggest residuals and refitting model 4.5 to the remaining 128 data points results in a residual sum of squares of 0.00249 (see Table 4.11) with 124 df. The pooled residual sum of squares, recalculated from

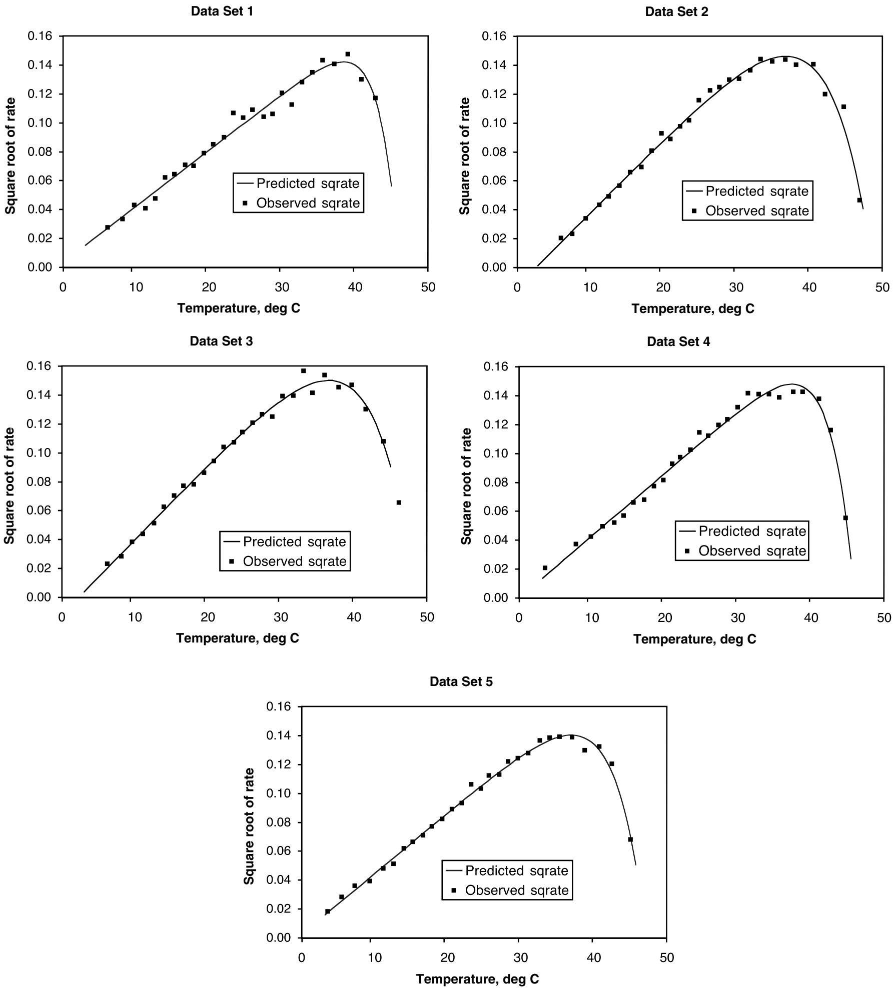


FIGURE 4.4 Five individual data sets and fitted square-root models. (Data from Nichols et al., *Appl. Environ. Microbiol.*, 68, 2809–2813, 2002.)

the five data sets with the outliers deleted, is 0.00134 with 108 df. The difference between these sums of squares is 0.00249 – 0.00134 = 0.00115 with 124 – 108 = 16 df. This leads to the following variance ratio test:

$$F_{16,108} = \frac{0.00115/16}{0.00134/108} = 5.79,$$

which is still highly significant ($P < 0.001$). From Table 4.11, we see that the P value for the normality test of the residuals is 0.445, indicating that the new set of residuals obtained by data elimination is close to having a normal distribution. Therefore, the significant variance ratio above cannot be attributed to outliers, but

TABLE 4.11
Results of Fitting Five Replicated Data Sets of Table A4.3

Data Set	n	Residual Sum of Squares	Residual Mean Square	P Value Test of Normality	No. of Runs of Residuals	Highest Hougaard Skewness	$b \pm$ SE	$T_{min} \pm$ SE	$c \pm$ SE	$T_{max} \pm$ SE
1	26	0.000661	0.000030	0.4837	15	0.956 (T_{max})	0.00388 ± 0.00016	–0.97 ± 0.91	0.3506 ± 0.0990	46.09 ± 0.99
2	27	0.000493	0.000021	0.0409	14	0.333 (T_{max})	0.00507 ± 0.00019	2.11 ± 0.59	0.1576 ± 0.0145	48.34 ± 0.25
3	27	0.000333	0.000014	0.0525	20	0.289 (T_{max})	0.00518 ± 0.00016	2.32 ± 0.48	0.1569 ± 0.0127	48.32 ± 0.23
4	27	0.000478	0.000021	0.8385	6	0.365 (T_{max})	0.00447 ± 0.00013	0.45 ± 0.60	0.2886 ± 0.0256	45.47 ± 0.17
5	28	0.000250	0.000010	0.1100	16	0.293 (T_{max})	0.00433 ± 0.00010	–0.23 ± 0.43	0.2192 ± 0.0163	46.32 ± 0.20
1–5	135	0.00677	0.000052	<0.0001	75	0.287 (T_{max})	0.00472 ± 0.00012	1.18 ± 0.43	0.1718 ± 0.0119	47.88 ± 0.22
1–5 (outliers deleted)	128	0.00249	0.000020	0.4447	74	0.200 (T_{max})	0.00474 ± 0.00008	1.22 ± 0.27	0.1678 ± 0.0077	48.20 ± 0.16

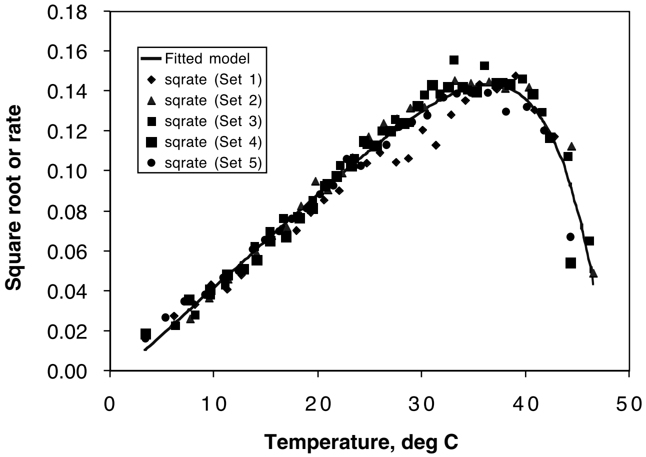


FIGURE 4.5 Pooled data and fitted overall square-root model (Data from Nichols et al., *Appl. Environ. Microbiol.*, 68, 2809–2813, 2002.)

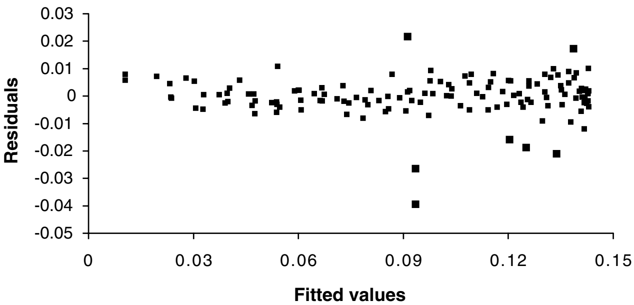


FIGURE 4.6 Residuals vs. fitted values from overall model (n = 135). The seven largest residuals are indicated using a larger font size. (Data from Nichols et al., *Appl. Environ. Microbiol.*, 68, 2809–2813, 2002.)

must be interpreted as indicating that there is significant variability among the five data sets.

Looking more closely at the five sets of data, we see from the dates given in Table A4.3 that Data Set 1 was obtained on 6 September 2000, Data Sets 2 and 3 were obtained on 25 September 2000, and Data Sets 4 and 5 were obtained on 14 December 2000. Data Sets 2 and 3 are, in fact, close replicates, as the "bar" run from which Data Set 2 was obtained used tubes on one side of the temperature gradient incubator, and the run from which Data Set 3 was obtained used tubes on the other side of the incubator. Different stock solutions were employed for the two runs, so that variation in the results would be a consequence of variation in the inocula and not in the ambient conditions of the room in which the incubator was housed. Similarly, Data Sets 4 and 5 are close replicates for the same reason.

We now compare Data Sets 2 and 3 using the same test we used above. From the results in Table 4.11, the pooled residual sum of squares is 0.000493 + 0.000333 = 0.000826 with 23 + 23 = 46 df. Combining the two data sets leads to a residual sum of squares of 0.000847 with 50 df. The difference between these sums of squares is 0.000847 – 0.000826 = 0.000021, with 50 – 46 = 4 df. This leads to the following variance ratio test:

$$F_{4,46} = \frac{0.000021 / 4}{0.000826 / 46} = 0.292$$

a nonsignificant F value that confirms that Data Sets 2 and 3 are close replicates.

Comparing Data Sets 4 and 5 in a similar way leads to a pooled residual sum of squares of 0.000728 with 47 df, while combining the two data sets leads to a residual sum of squares of 0.00103 with 51 df. The difference between these sums of squares is 0.00103 – 0.000728 = 0.000302 with 51 – 47 = 4 df, leading to the following variance ratio test:

$$F_{4,47} = \frac{0.000302 / 4}{0.000728 / 47} = 4.90$$

This variance ratio is significant ($P < 0.005$), so the two data sets, although obtained on the same bar run, cannot be considered to be close replicates. Removing the data point responsible for the largest residual (T = 38.1°C; sqrate = 0.1297; Data Set 5) still results in a significant variance ratio ($F_{4,46} = 3.49$), so the lack of closeness of agreement of these two data sets cannot be attributed to a single outlying data point. From Figure 4.5, one can visually observe that there are differences between Data Set 4 and Data Sets 2 and 3. A formal test leads to $F_{8,69} = 17.26$, which is highly significant. Hence, one must conclude that the data sets obtained on 14 December are different from those obtained on 25 September. All are different from Data Set 1, obtained on 6 September, which exhibited a series of low rates in the mid-temperature range.

Identifying deviant data points using analysis of residuals is desirable as a means of directing attention to the possibility of experimental errors. However, one should avoid deleting data points with large residuals unless there are good, objective reasons for doing so, because of belief that errors were committed during the experiment, resulting in erroneous readings. The indiscriminate use of data elimination may have the undesirable effect of leading to biased estimates of the parameters, and may produce a belief that the data set, modified by having its most deviant points deleted, is of a better quality than is really justified.

Because five replicate data sets for the growth of *L. monocytogenes* were available, it was possible to see that, although each data set looked to be good in isolation, the aberrant nature of Data Set 1 became apparent when all data were pooled and plotted on a single graph (Figure 4.5). To the author's knowledge, very few replicate data sets are available for turbidity measurements obtained using a temperature

gradient incubator. Replication is desirable as it makes it possible to examine whether the variation in experimental data is significant or not.

What explanations might be offered to explain the variation among the five data sets discussed here? An obvious one is operator error in using the temperature gradient incubator, but this is unlikely in the present case because the experimenters were very experienced. Another explanation has to do with their use of the modified Gompertz function as a primary model to estimate the lag time and the maximum specific growth rate at each temperature. That model may not have been an appropriate one for determining these parameters, especially as the cultures were harvested when the individual incubation tubes reached a transmittance value of only 27 to 30%. A second explanation may be that the secondary model, the four-parameter square-root model (4.5), was inadequate or inappropriate. This is unlikely, as the information in Figure 4.4 and in Table 4.11 indicates that 4.5 is a good model for these data. Another explanation may lie with the equipment itself, or with the room in which the incubator was housed, as its ambient temperature is probably not adequately controlled. Although it is not possible to be certain about the true explanation for the variation observed among the data sets in question, it is always important for investigators to consider the various possibilities, so that one may improve the experimental procedure, or the modeling process that follows collection of the data, or both.

4.3 UNCERTAINTY IN LAG TIMES AND GENERATION TIMES, AND ITS CONSEQUENCES

In this section, we look at uncertainty in some of the basic parameters that are measured or derived from experimentation and data analysis in predictive food microbiology, and its consequences for food production and safety. Uncertainty is an ever-present phenomenon, one which may be reduced by careful quality control, but which may never be totally eliminated.

A fundamental issue of the discipline of predictive food microbiology (perhaps its most important one) is the accurate prediction of the shelf life of a food product so that the product remains edible throughout the whole of that period. We have seen in some of the earlier sections of this chapter (e.g., 4.2.4.2) that modeling of some of the important derived parameters such as the maximum specific growth rate μ occurs in a transformed rate domain; that is, μ, which has units of reciprocal time, is transformed by taking its square root (e.g., Equation 4.5 or Equation 4.6a) or its logarithm (e.g., Equation 4.5b or Equation 4.6b) when it is to be used in a modeling exercise or various other statistical calculations. The reason for applying one of these transformations is that the untransformed growth rate has undesirable statistical properties. Generally, the probability distribution of μ is nonnormal (i.e., not a Gaussian distribution), which means, among other things, that the variance of the distribution is not independent of its mean. By applying a transformation such as the square root or logarithm to the growth rate, the transformed variable comes close to having a normal distribution. The random variable will then be symmetrically distributed around the mean, with a variance that does not depend upon the mean. Since the mean

and variance are the two parameters of the normal distribution, having a good estimate of these parameters from a sample of size n (say) enables the user to have confidence about the probability of obtaining an observation from the distribution of the transformed variable that falls outside of any specified bound. For example, the probability of an observation being more than two standard deviations from the mean can be calculated using a table of the t distribution with $n - 1$ df. If the random variable is not normally distributed, one may be able to make a similar calculation if its distribution is known, but usually with more difficulty than if the random variable is normal.

The situation is even more complicated when one considers time-based variables or derived parameters such as lag time, generation time, and shelf life. While square roots of rate or logarithms of rate may be normally distributed, lag times, generation times, and shelf lives never are, tending instead to have long-tailed distributions. In the following subsection, we review the distributional features of these time-domain quantities.

4.3.1 The Distribution of Lag Time and Generation Time

The distribution of lag time and generation time is a nonnormal distribution of which the variance usually falls somewhere between being proportional to the square of the mean or to the cube of the mean; that is, if we denote the mean (lag or generation) time by θ, then the variance, if denoted by V, is usually given by

$$V = c\theta^2 \tag{4.14}$$

or

$$V = c\theta^3 \tag{4.15}$$

or by an exponent of θ that is a noninteger value lying somewhere between 2 and 3, with c being a proportionality constant (see McMeekin et al., 1993; Ratkowsky et al., 1991; Schaffner, 1998). There are some well-known probability distributions with variances having the properties defined above. For example, the gamma and Weibull distributions have the variance proportional to the square of the mean, as given by 4.14, and the inverse-Gaussian distribution has its variance proportional to the cube of the mean, as given by 4.15. There are other lesser-known distributions with the same properties.

Data to test whether the exponent of θ is 2, 3, or some other value are hard to find, as they require many replicate runs carried out at each of a sequence of temperatures. Although one might expect to obtain a reasonable prediction of the population mean from a relatively small sample size, say 10 to 15 values, sample variances are more variable than sample means and many dozens of trials are needed to obtain a good prediction of the variance. Using the data of Neumeyer given in Appendix 4A.4 of McMeekin et al. (1993) on generation times for the growth of *Staphylococcus aureus* 3b from 12.5 to 35C, it was established that 4.15 was a reasonable model for the relationship between the variance and the mean (see Table 4.8 and associated text of McMeekin et al. [1993] for the methodology used).

Nevertheless, sample sizes were rather small, ranging from $n = 1$ to 18. Only two temperatures had $n > 10$.

A far larger data set ($n = 125$) for examining the question of the correct model between a probability distributions variance and mean is that of Macario, reported in Ratkowsky et al. (1996), for the generation time of *Pseudomonas fluorescens* in the temperature range of 2.4 to 16.3°C, obtained using nutrient broth in a temperature gradient incubator. Grouping the temperature data into 1°C intervals produced 15 intervals with sample sizes ranging between 2 and 17. The estimates of the means and variances were plotted against temperature as the ratios V/θ, V/θ^2, and V/θ^3. The regression line of V/θ^2 against temperature was the one with the least correlation, suggesting that 4.14 was the best model for those data. This suggested to Ratkowsky et al. (1996) that the Macario data could be modeled by 4.14 with the scale parameter $c = 0.006676$ estimated by assuming that a gamma distribution was a suitable probability distribution for the data (see Ratkowsky et al., 1996, for a detailed description of the methodology employed). It should have been realized that the gamma distribution, which contains only two parameters, is only one possible probability distribution having the property given by 4.14. Although the assumption of a gamma distribution led to predicted variances that were in the right "ball park" when compared with the experimental variances, the predicted standardized skewness coefficient of 0.163 indicated only a small amount of skewness, as the histograms in Figure 2 of Ratkowsky et al. (1996) were scarcely distinguishable from those of a normal distribution. Later, Hutchinson (1998) pointed out that the error made by Ratkowsky et al. (1996) was to believe that V/θ^2 being a constant told one something about the shape of the distribution, e.g., whether it was skewed.

Hutchinson (1998) pointed out that to determine the real shape of the distribution, one would have to look at the data for each temperature separately, and that would require dozens of observations at each temperature. Clearly, the Macario data set, averaging ca. 8 points per temperature, was too small for this. He further showed that if one could assume that the "shifted gamma distribution" were a suitable probability distribution for those data, having a third parameter that measured the extent to which the mean was shifted upwards from zero, one could obtain a fitted distribution with a considerable skewness. However, until such time as microbiologists can readily produce many replicated data sets at each temperature, the question of the true amount of skewness in data that obey models 4.14 and 4.15 will not be answered. In the following section, we explore how knowledge of the true distribution for lag time and generation time may be used to obtain accurate predictions about the likely shelf life of a food product.

4.3.2 The Prediction of Shelf Life

The shelf life (keeping time of a food product), subject to spoilage owing to a spoilage organism, can be predicted from lag time and generation time if there are appropriate models for these parameters expressed in terms of temperature T, water activity a_w, hydrogen ion concentration (pH), concentration of undissociated weak acid, and any other factors influencing them.

Denoting mean lag time by t_L and mean generation time by t_G, the shelf life may be estimated by the expression

$$\text{Shelf life} = t_L + t_G \ln(C_f/C_0)/\ln 2 \tag{4.16}$$

where C_0 is the initial bacterial concentration and C_f is the maximum permissible concentration (e.g., in cfu/ml), where permissible means the product is deemed to be spoiled. The first term of the expression is the lag time before growth effectively begins, and the longer the value of t_L, the longer the shelf life. The second term of the expression calculates the number of generations through which the microorganism develops from its initial concentration to its final concentration. Note that the above assumes that the generation time is independent of the numbers of bacteria present. If that assumption is too naïve, or too crude an approximation, one may predict C_f from a model (such as the modified Gompertz curve — see chapter on Primary Modeling).

An example of the use of 4.16 will now be presented. Consider the data on the growth of *Aeromonas hydrophila* (aerobic atmosphere), from Palumbo et al. (1991). For T = 19°C, added salt = 3.5%, pH = 6.3 and added sodium nitrite = 50, and assuming $C_0 = 10$ and $C_f = 1.0 \times 10^7$ cfu/ml, the lag time t_L was determined to be 60 h and the generation time t_G was determined to be 2.6 h, using a square-root model. Using 4.16, the keeping time is calculated as

$$\text{Shelf life} = 60.0 + 2.6 \ln(10^7/10)/\ln 2 = 112 \text{ h}$$

The second approach, which does not assume a constant generation time, can be made using a computer package such as the pathogen modeling program (PMP), developed by the USDA/ARS/ERRC Microbial Food Safety Unit. From PMP, the lag phase duration t_L is 48.2 and the generation time t_G is 2.7 h, corresponding to the above environmental factors. The calculated time to reach the level of concern 10^7 cfu/ml is 105 h, which is similar to the 112 h calculated using 4.16.

The important thing about the above calculation is that in either approach only mean values of lag or generation times are used. The variability of these parameters, and their probability distributions, as discussed in Section 4.3.1, has not entered into the calculation. We will now take a brief look at how variability may affect the result obtained, and lead to a larger or smaller shelf life.

We reconsider the data of Macario, discussed in Section 4.3.1. It was found that it was reasonable to assume that the ratio of the variance to the square of the mean was constant, which is equivalent to assuming that the coefficient of variation is constant. Ratkowsky et al. (1996) further assumed, naively as Hutchinson (1998) later demonstrated, that this implied that one could use the gamma distribution, which has two parameters, for these data. Estimating the scale parameter to be c = 0.006676, we calculated predicted mean generation times for any temperature and the distribution of generation time under the assumption that the gamma distribution was an appropriate one for those data. For example, at T = 2.4C, the mean generation time was determined to be θ = 615.2 min (see Table 2 of Ratkowsky et al., 1996) and a series of predicted probabilities, denoted by θ_0, were calculated at that

temperature. Thus, for $P = 0.000001$, $\theta_0 = 405.1$ min, for $P = 0.001$, $\theta_0 = 471.5$ min, for $P = 0.999$, $\theta_0 = 782.3$ min, for $P = 0.999999$, $\theta_0 = 885.8$ min; and so on (see Table 4 of Ratkowsky et al., 1996). What do these numbers signify? For example, they tell us that although the *mean* generation time may be 615.2, there is a one in a thousand chance that the generation time may be as low as 471.5 and a one in a million chance that it may be as low as 405.1. From the point of view of food acceptability, these values, if used in 4.16 in place of 615.2, would lead to a much-reduced shelf life. At the other end of the scale, there is a one in a thousand chance that the generation time may be as high as 782.3 and a one in a million chance that it may be as high as 885.8. These values would lead to a prolonged shelf life compared to the expected shelf life based upon the mean generation time. Note that 471.5 and 782.3 are not equally distant from 615.2, and that neither are 405.1 and 885.8. This is a consequence of the asymmetry inherent in a distribution such as the gamma distribution. However, as pointed out by Hutchinson (1998), this distribution is not particularly skewed, and to get a good estimate of the true skewness would require many replicates at each temperature.

4.4 EPILOGUE

In this epilogue, the author raises a few issues that he has considered over the years. The first two subsections deal with beliefs, held by at least some modelers, that the author feels are erroneous. Because they have appeared in the food microbiology literature, these issues need to be raised. In the final two subsections, some newer or less familiar modeling methods are discussed, which have found, or will find, their way into the predictive modeling literature in the near future.

4.4.1 Use of the Expressions "Fail Safe" and "Fail Dangerous" for Models

There seems to be a widespread use of the expressions "fail safe" and "fail dangerous" in the food microbiological literature when applied to models. Assuming that one has unbiased, carefully collected data, fail-safe refers to a model that overpredicts the rate at which a spoilage or pathogenic organism will grow, or predicts overly stringent conditions at the growth/no growth interface for growth to occur. Conversely, fail-dangerous refers to a model that under-predicts the actual growth rate or fails to predict conditions at which growth will actually occur. In either case, the model must be deemed to be inadequate. To use the expression fail-safe to exonerate, exculpate, or absolve the modeler, scientist, or regulator from doing a better job seems to be just taking the easy way out. There is really only one kind of model that should find a place in the food microbiology literature, and that is a *good* model. Good models are those that closely mimic the rate at which spoilage or pathogenic organisms grow or which closely predict the position of the growth/no growth interface. Bad models are all other kinds.

The issue is not an academic one, but a practical one. Fail-safe can be taken to a ludicrous extreme, by adopting a "model" that always predicts that a pathogenic microorganism will grow, even under conditions so stringent as to be virtually

impossible. Using this model, all food products would be declared unsafe to consume. We live on a planet in which a very high proportion of its human inhabitants do not get enough food to eat. Death by starvation is still a global problem and adequate nutrition, to help people ward off debilitating or potentially fatal diseases, is a worldwide problem. An extreme fail-safe model might suggest that food that is in reality safe to eat should be avoided or destroyed. We should always remember that death and disease could be caused as much by the nonavailability of food as well as by its being contaminated or spoiled. In the more affluent world economies, a fail-safe mentality also overlooks the producer's point of view. Food that is safe to be sold and consumed may, as a result of overzealous regulations resulting from inadequate modeling, have to be discarded, leading not only to lower profitability but less availability and choice to the consumer. In addition, any resulting added costs tend to get passed on to the consumer.

4.4.2 Correlation between Parameters

There seems to be a strong belief on the part of certain researchers that a regression model with low correlation between its parameters is, in some sense, "better" than one with high parameter correlation. For example, Rosso et al. (1993) found that the three cardinal temperature parameters in 4.6 were linearly correlated, a condition that they felt was "unexpected" (see title of their paper). No rational reasons were advanced as to why uncorrelated parameters are deemed to be superior to correlated ones. It is the present author's experience that there is no connection between parameter correlation and the properties of the estimators of the parameters in nonlinear regression models (Ratkowsky, 1983, 1990). It simply has nothing to do with the more important question of whether the estimators of the parameters are unbiased, jointly normally distributed, and attain the so-called "minimum variance bound." It might be nice to have a nonlinear regression model that not only is "close to linear" (see Section 4.2.5.4), but also has uncorrelated parameters. The present writer has never found such a model.

4.4.3 Artificial Neural Networks as an Alternative to Statistical Modeling

Various alternatives to statistical modeling are starting to appear in the predictive microbiology literature, some of which are likely to prove appealing to a new generation of modelers. Probably the most important of these involves the use of artificial neural networks (ANN) in one form or other (they have also been referred to as computational neural networks, artificial computational neural networks, or general regression neural networks). An early exposition in the predictive microbiology literature of the methodology of this approach, which would be classified as a black box approach using the system of Ljung (1999), is that given by Hajmeer et al. (1997), who described some of the computational details and gave an example of its use on microbial growth data. ANNs operate by analogy to the human nervous system, where input variables provide an incoming signal to a neuron, are modified in a "hidden" neuron layer, and finally converted to an output signal by an appropriate

"transfer function." The application by Hajmeer et al. (1997) to anaerobic growth data obtained by Zaika et al. (1994) on *Shigella flexneri* as a function of pH, NaCl, and $NaNO_2$ concentrations, used a hidden layer consisting of 20 nodes (neurons), because their "training cycles" indicated that their goodness-of-fit measures became stable when 20 nodes were used. Some compromises need to be made, since by increasing the number of training cycles and the number of hidden nodes indefinitely, one can fit "training sets" perfectly, but at the risk of poorly predicting subsequent validation (testing) sets. This is, of course, directly analogous to regression modeling, where the use of too many terms in a regression model may result in overfitting the original data set and a poor fit to validation data sets.

Geeraerd et al. (1998) used a low-complexity ANN to convert the incoming signal from environmental factors such as temperature, pH and salt concentration to an output signal embodied in parameters such as the maximum growth rate, the lag time, and the initial population size. They referred to their model as a "hybrid gray box" model because the ANN models were used only to describe the effect of an influencing factor such as temperature on an output parameter such as μ_{max}, which was then used in a dynamic growth model to predict the growth of the microbial population with time.

Jeyamkondan et al. (2001) used commercially available neural network software, which enabled them to examine different network structures quickly with little effort, a feature that is appealing to users who know some basic principles of neural networks but are not experts in neural network programming. They chose a general regression neural network (GRNN) to predict response parameters such as generation time and lag phase duration from input data involving changes in temperature, NaCl, pH, etc. for three microorganisms. They compared use of the GRNN to that of more traditional statistical models and found that the GRNN predictions were far superior to predictions from statistical models for training data sets, but similar to, or slightly worse than statistical model predictions for test data sets. They concluded that neural networks were adequate for food safety tests and for new product development. To assess goodness-of-fit, they investigated the performance of various statistical indices and concluded that the use of the mean absolute relative residual, the mean relative percentage residual, and the root mean square residual, in conjunction with graphical plots such as the bias plot and the residual plot, were sufficient for comparing competing models.

More recently, Hajmeer and Basheer (2002, 2003) proposed a probabilistic neural network (PNN) for use on growth/no growth data and compared its classificatory performance to that of a linear logistic regression model, a nonlinear logistic regression model of the kind proposed by Ratkowsky and Ross (1995), and a feedforward error backpropagation artificial neural network (FEBANN), and found that the optimal PNN gave the lowest misclassification rates. It should be pointed out, though, that the "nonlinear" logistic regression model that they derived using their training data subset was not a true nonlinear logistic model, since the cardinal parameters $a_{w\,min}$, T_{min}, and T_{max} were assumed to be fixed constants, taken from the paper of Salter et al. (2000), and not estimated as free parameters. Because of this, the resulting model was really a linear logistic model and not capable of doing as well as a true nonlinear logistic regression model.

At this early stage in the use of ANNs, it is difficult to forecast how valuable they might be in predictive microbiology to predict the conditions under which food products should be stored to guarantee their safety and quality. One question that arises is whether ANN models are "portable," i.e., whether other workers can readily use them in the way that they can use statistical models that have expressions in the form of mathematical equations. In this regard, hybrid models may prove to be attractive, where the ANN may serve to produce a primary model, in a similar way in which the modified Gompertz model or the model of Baranyi et al. (1993) is used, followed by a more traditional secondary model where the outputs from the ANN are then expressed as functions of the environmental factors.

4.4.4 Principal Components Regression and Partial Least-Squares Regression

There are other alternatives to ANN that might be used by a new generation of modelers. Jeyamkondan et al. (2001) mentioned principal component regression (PCR) and partial least-squares regression (PLS) as two statistical techniques of multivariate analysis that might be employed when the underlying relationships are not known. PCR has been known for some time, and is available in many standard statistical packages. Given a set of n experimental units on which measurements have been made of p explanatory variables, a principal component analysis can reduce those p variables to a smaller set (say 2 or 3) of "canonical" variates, upon which regression analysis of various response variables may then be performed. The canonical variates (i.e., the principal components) are linear combinations of the p explanatory variables and have the property that they are orthogonal (i.e., uncorrelated) to each other, unlike the original set of p variables, at least some of which are likely to be highly correlated. If the canonical variates are easily interpretable, the contributions of each variate to the explanation of the response variable can be quantified, because of their orthogonality. Hence, PCR has the potential to be a useful technique, provided that the canonical variates are subject to interpretation.

PLS is a much newer technique and at the moment is poorly understood by users, but is gaining increasing application. It can be contrasted with PCR and with another technique, called reduced rank regression (RRR), when the response variables form a multivariate set. Whereas PCR extracts successive linear combinations of the explanatory variables to explain as much *predictor* sample variation as possible, RRR extracts successive linear combinations of the set of response variables to explain as much *response* sample variation as possible. PLS tries to balance the two objectives by simultaneously explaining response variation and predictor variation. The same caveats that applied to ANN modeling apply here as well. Just as the use of too many nodes or too many training cycles can lead to over-fitting the training set of data and poor prediction in test data sets for ANN modeling, extracting too many factors in PLS can also lead to overfitting. Like ANN modeling, the use of validation, by splitting one's data set into a training set and a test set, is an integral part of the modeling process.

Time can only tell how useful such techniques might be to predictive microbiology, but one has to anticipate that a new generation of modelers is certain to come forward with applications employing one or more of these procedures. One must retain an open mind to their use, but at the same time avoid uncritical acceptance of them. In addition, the restriction inherent in all these techniques, which involve *linear* combinations of variables, may limit the general applicability of the methodology. After all, the world we live in is not a linear one, and it is a rare circumstance in mathematical modeling when a linear model explains natural phenomena adequately.

APPENDIX

TABLE A4.1
Data Set on Lag Time

	Lag Time (hr)	
***T*(°C)**	**Breast**	**Thigh**
8	43.8	46.8
10	19.6	21.6
12	14.9	10.3
14	11.3	9.1
16	6.5	5.7
18	5.3	4.5
20	3.7	3.9
22	3.8	3.2
24	3.3	2.8
26	2.5	2.4
28	2.2	2.2
30	2.1	2.3
32	1.6	1.5
34	1.4	1.6
36	1.4	1.4
38	1.3	1.4
40	1.3	1.2
42	1.0	1.5
44	1.1	1.1
46	0.9	1.0
48	1.6	1.0

Source: From Oscar, T.P., *Int. J. Food Microbiol.*, 76, 177–190, 2002. With permission.

TABLE A4.2
Specific Growth Rate Constant μ vs. Temperature for Three Data Sets

Alteromonas sp. (CLD38)		*Pseudomonas* sp. (16L16)		*Morganella morganii* (M68)	
T (°C)	μ	*T* (°C)	μ	*T* (°C)	μ
1.3	0.00597	0.0	0.00588	19.0	0.002032
2.8	0.00805	1.5	0.00854	19.8	0.002320
4.0	0.01036	2.4	0.01083	21.5	0.002660
5.2	0.01269	4.1	0.01376	23.5	0.002924
7.2	0.01742	5.2	0.01634	24.7	0.003378
7.9	0.02141	7.1	0.02179	25.6	0.003759
9.2	0.02315	7.9	0.02667	26.8	0.004000
11.4	0.03049	9.3	0.02896	28.2	0.004273
12.4	0.03226	11.3	0.03810	29.3	0.004717
17.2	0.05285	13.3	0.04577	30.3	0.005181
18.3	0.05656	14.3	0.05157	31.6	0.005464
19.5	0.05960	17.4	0.06925	33.0	0.005848
21.7	0.06757	18.5	0.07519	34.5	0.005917
23.0	0.07052	19.5	0.07752	36.0	0.006098
24.3	0.07364	22.0	0.09728	37.4	0.006098
25.6	0.07407	23.1	0.1015	38.8	0.006250
28.1	0.06849	24.5	0.1111	40.0	0.005181
29.9	0.06075	25.6	0.1155	41.5	0.003623
		28.1	0.1256		
		29.9	0.1230		
		31.6	0.1171		

TABLE A4.3
Growth Rate Data of *Listeria monocytogenes*, Presented as Square Root of Rate vs. Temperature, Five Replicates

Data Set 1 (6 Sept 2000)		Data Set 2 (25 Sept 2000)		Data Set 3 (25 Sept 2000)		Data Set 4 (14 Dec 2000)		Data Set 5 (14 Dec 2000)	
T (°C)	$\sqrt{\mu}$	*T* (°C)	$\sqrt{\mu}$	*T* (°C)	$\sqrt{\mu}$	*T* (°C)	$\sqrt{\mu}$	*T* (°C)	$\sqrt{\mu}$
1.8	0	6.2	0.0232	6.2	0.0229	0.7	0	0.8	0
3.8	0	7.7	0.0262	8.1	0.0280	3.4	0.0185	3.4	0.0163
6.1	0.0277	9.5	0.0365	9.6	0.0378	7.6	0.0356	5.3	0.0266
8.2	0.0335	11.3	0.0459	11.1	0.0434	9.6	0.0407	7.1	0.0345
9.7	0.0432	12.6	0.0516	12.6	0.0509	11.2	0.0480	9.1	0.0379
11.3	0.0411	14.1	0.0591	13.9	0.0622	12.8	0.0506	10.9	0.0465
12.6	0.0478	15.5	0.0681	15.4	0.0697	14.1	0.0558	12.3	0.0500
14.0	0.0623	17.0	0.0719	16.7	0.0765	15.4	0.0650	13.7	0.0607
15.3	0.0646	18.4	0.0827	18.0	0.0776	16.9	0.0672	14.9	0.0654
16.7	0.0712	19.7	0.0948	19.5	0.0856	18.2	0.0766	16.3	0.0700
17.9	0.0703	21.0	0.0909	20.8	0.0938	19.5	0.0811	17.5	0.0761
19.3	0.0792	22.3	0.0994	22.1	0.1032	20.7	0.0927	18.9	0.0814
20.6	0.0854	23.5	0.1036	23.5	0.1065	21.8	0.0971	20.2	0.0884
22.0	0.0902	24.8	0.1173	24.7	0.1137	23.2	0.1024	21.5	0.0926
23.3	0.1071	26.3	0.1240	26.1	0.1201	24.4	0.1150	22.8	0.1059
24.7	0.1039	27.6	0.1262	27.4	0.1257	25.6	0.1125	24.1	0.1027
26.0	0.1092	28.9	0.1314	28.8	0.1243	27.0	0.1201	25.2	0.1121
27.5	0.1045	30.2	0.1320	30.2	0.1384	28.2	0.1240	26.6	0.1128
28.7	0.1064	31.8	0.1375	31.7	0.1388	29.6	0.1326	27.8	0.1218
30.0	0.1207	33.2	0.1455	33.1	0.1559	31.0	0.1429	29.1	0.1242
31.3	0.1129	34.7	0.1437	34.3	0.1407	32.5	0.1419	30.5	0.1278
32.8	0.1284	36.5	0.1449	36.0	0.1530	33.9	0.1421	32.1	0.1368
34.2	0.1353	38.0	0.1414	37.9	0.1445	35.3	0.1398	33.4	0.1386
35.6	0.1435	40.3	0.1419	39.7	0.1462	37.2	0.1437	34.7	0.1393
37.2	0.1410	41.9	0.1214	41.6	0.1296	38.5	0.1436	36.4	0.1391
39.1	0.1478	44.4	0.1128	44.0	0.1071	40.7	0.1389	38.1	0.1297
40.9	0.1304	46.6	0.0490	46.2	0.0650	42.3	0.1166	40.1	0.1323
42.8	0.1175					44.3	0.0540	41.8	0.1201
45.0	0					46.5	0	44.3	0.0671
47.1	0							46.3	0

REFERENCES

Adams, M.R., Little, C.L., and Easter, M.C. (1991). Modelling the effect of pH, acidulant and temperature on the growth rate of *Yersinia enterocolitica. J. Appl. Bacteriol.* 71: 65–71.

Baranyi, J., Roberts, T.A., and McClure, P.J. (1993). A non-autonomous differential equation to model bacterial growth. *Food Microbiol.* 10: 43–59.

Baranyi, J., Ross, T., McMeekin, T.A., and Roberts, T.A. (1996). Effects of parameterization on the performance of empirical models used in predictive microbiology. *Food Microbiol.* 13: 83–91.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Cole, M.B., Jones, M.V., and Holyoak, C. (1990). The effect of pH, salt concentration and temperature on the survival and growth of *Listeria monocytogenes. J. Appl. Bacteriol.* 69: 63–72.

Cook, R.D. (1979). Influential observations in linear regression. *J. Am. Statist. Assoc.* 74:169–174.

Dantigny, P. and Molin, P. (2000). Influence of the modeling approach on the estimation of the minimum temperature for growth in Bělehrádek-type models. *Food Microbiol.* 15, 185–189.

Davey, K.R. (1989). A predictive model for combined temperature and water activity on microbial growth during the growth phase. *J. Appl. Bact.* 67: 483–488.

Davey, K.R. (1991). Applicability of the Davey (linear Arrhenius) predictive model to the lag phase of microbial growth. *J. Appl. Bact.* 70: 253–257.

Davey, K.R. and Amos, S.A. (2002). Letter to the editor. *J. Appl. Microbiol.* 92:583–584.

Delignette-Muller, M.L. (1998). Relation between the generation time and the lag time of bacterial growth kinetics. *Int. J. Food Microbiol.* 43: 97–104.

Fox, J. (1991). *Regression Diagnostics* (Sage University Paper series on quantitative applications in the social sciences, Series No. 07-079). Newbury Park, CA: Sage.

Geeraerd, A.H., Herremans, C.H., Cenens, C., and Van Impe, J.F. (1998). Application of artificial neural networks as a non-linear modular modeling technique to describe bacterial growth in chilled food products. *Int. J. Food Microbiol.* 44: 49–68.

Geeraerd, A.H., Valdramidis, V.P., Devlieghere, F., Bernaerts, H., Debevere, J., and Van Impe, J.F. (2002). Development of a novel modelling methodology by incorporating *a priori* microbiological knowledge in a black box approach. In *Proceedings and Abstracts of the 18th International ICFMH Symposium, Food Micro 2002*, Lillehammer, Norway, 18–23 August 2002, Axelsson, L., Tronrud, E.S., and Merok, K.J. (Eds.), pp. 135–138.

Hajmeer, M. and Basheer, I. (2002). A probabilistic neural network approach for modeling and classification of bacterial growth/no-growth data. *J. Microbiol. Methods* 51: 217–226.

Hajmeer, M. and Basheer, I. (2003). Comparison of logistic regression and neural network-based classifiers for bacterial growth. *Food Microbiol.* 20: 43–55.

Hajmeer, M.N., Basheer, I.A., and Najjar, Y.M. (1997). Computational neural networks for predictive microbiology II. Application to microbial growth. *Int. J. Food Microbiol.* 34: 51–66.

Houtsma, P.C., Kant-Muermans, M.L., Rombouts, F.M., and Zwietering, M.H. (1996). Model for the combined effects of temperature, pH and sodium lactate on growth rates of *Listeria innocua* in broth and Bologna-type sausages. *Appl. Environ. Microbiol.* 62: 1616–1622.

Hutchinson, T.P. (1998). Note on probability distributions for generation time. Letter to the editor. *J. Appl. Microbiol.* 85: 192–193.

Jeyamkondan, S., Jayas, D.S., and Holley, R.A. (2001). Microbial growth modelling with artificial neural networks. *Int. J. Food Microbiol.* 64: 343–354.

Ljung, L. (1999). *System Identification Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Lowry, R.K. and Ratkowsky, D.A. (1983). A note on models for poikilotherm development. *J. Theor. Biol.* 105: 453–459.

McMeekin, T.A., Chandler, R.E., Doe, P.E., Garland, C.D., Olley, J., Putro, S., and Ratkowsky, D.A. (1987). Model for the combined effect of temperature and water activity on the growth rate of *Staphylococcus xylosus. J. Appl. Bacteriol.* 62: 543–550.

McMeekin, T.A., Olley, J.N., Ross, T., and Ratkowsky, D.A. (1993). *Predictive Microbiology: Theory and Application*. Taunton, Somerset: Research Studies Press.

Mendenhall, W. and Sincich, T. (1996). *A Second Course in Statistics: Regression Analysis*, 5th ed. Upper Saddle River, NJ: Prentice-Hall.

Nagelkerke, N.J.D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* 78: 691–692.

Nichols, D.S., Presser, K.A., Olley, J., Ross, T., and McMeekin, T.A. (2002). Variation of branched-chain fatty acids marks the normal physiological range for growth in *Listeria monocytogenes*. *Appl. Environ. Microbiol.* 68: 2809–2813.

Oscar, T.P. (2002). Development and validation of a tertiary simulation model for predicting the potential growth of *Salmonella typhimurium* on cooked chicken. *Int. J. Food Microbiol.* 76: 177–190.

Palumbo, S.A., Williams, A.C., Buchanan, R.L., and Phillips, J.G. (1991). Model for the aerobic growth of *Aeromonas hydrophila* K144. *J. Food Protection* 54: 429–435.

Pin, C. and Baranyi, J. (1998). Predictive models as a means to quantify the interactions of spoilage organisms. *Int. J. Food Microbiol.* 41: 59–72.

Presser, K.A., Ratkowsky, D.A., and Ross, T. (1997). Modelling the growth rate of *Escherichia coli* as a function of pH and lactic acid concentration. *Appl. Environ. Microbiol.* 63: 2355–2360.

Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach.* New York: Marcel Dekker.

Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. New York: Marcel Dekker.

Ratkowsky, D.A. and Ross, T. (1995). Modelling the bacterial growth/no growth interface. *Lett. Appl. Microbiol.* 20: 29–33.

Ratkowsky, D.A., Olley, J., McMeekin, T.A., and Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. *J. Bacteriol*. 149: 1–5.

Ratkowsky, D.A., Lowry, R.K., McMeekin, T.A., Stokes, A.N., and Chandler, R.E. (1983). Model for bacterial growth rate throughout the entire biokinetic temperature range. *J. Bacteriol*. 154: 1222–1226.

Ratkowsky, D.A., Ross, T., McMeekin, T.A., and Olley, J.N. (1991). Comparison of Arrhenius-type and Bělehrádek-type models for the prediction of bacterial growth in foods. *J. Appl. Bacteriol.* 71: 452–459.

Ratkowsky, D.A., Ross, T., Macario, N., Dommett, T.W., and Kamperman, L. (1996). Choosing probability distributions for modeling generation time variability. *J. Appl. Bacteriol.* 80: 131–137.

Rosso, L., Lobry, J.R., and Flandrois, J.P. (1993). An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. *J. Theor. Biol.* 162: 447–463.

Salter, M.A., Ratkowsky, D.A., Ross, T., and McMeekin, T.A. (2000). Modelling the combined temperature and salt (NaCl) limits for growth of a pathogenic *Escherichia coli* strain using nonlinear logistic regression. *Int. J. Food Microbiol.* 61: 159–167.

SAS Institute Inc. (1990). *SAS/STAT User's Guide*, Version 6, 4th ed. Cary, NC: SAS Institute Inc. (The SAS/STAT User's Guide for more recent releases of SAS such as Version 8 are available as on-line documentation by connecting to the SAS website.)

Schaffner, D.W. (1998). Predictive food microbiology Gedanken experiment: why do microbial growth data require a transformation? *Food Microbiol.* 15: 185–189.

Schoolfield, R.M., Sharpe, P.J.H., and Magnuson, C.E. (1981). Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. *J. Theor. Biol.* 88:719–731.

Sen, A. and Srivastava, M. (1990). *Regression Analysis: Theory, Methods, and Applications*. New York: Springer-Verlag.

Sharpe, P.J.H. and DeMichele, D.W. (1977). Reaction kinetics and poikilotherm development. *J. Theor. Biol.* 64:649–670.

Van Impe, J.F., Bernaerts, K., Geeraerd, A.H., Poschet, F., and Versyck, K.J. (2001). Modelling and prediction in an uncertain environment. In *Food Process Modelling*, Tijskens, L.M.M., Hertog, M.L.A.T.M., and Nicolai, B.M. (Eds.), Woodhead Publishing, Cambridge, U.K., pp.156–179.

Zaika, L.L., Moulden, E., Weimer, L., Phillips, J.G., and Buchanan, R.L. (1994). Model for the combined effects of temperature, initial pH, sodium chloride and sodium nitrite concentrations on anaerobic growth of *Shigella flexneri*. *Int. J. Food Microbiol.* 23: 345–358.